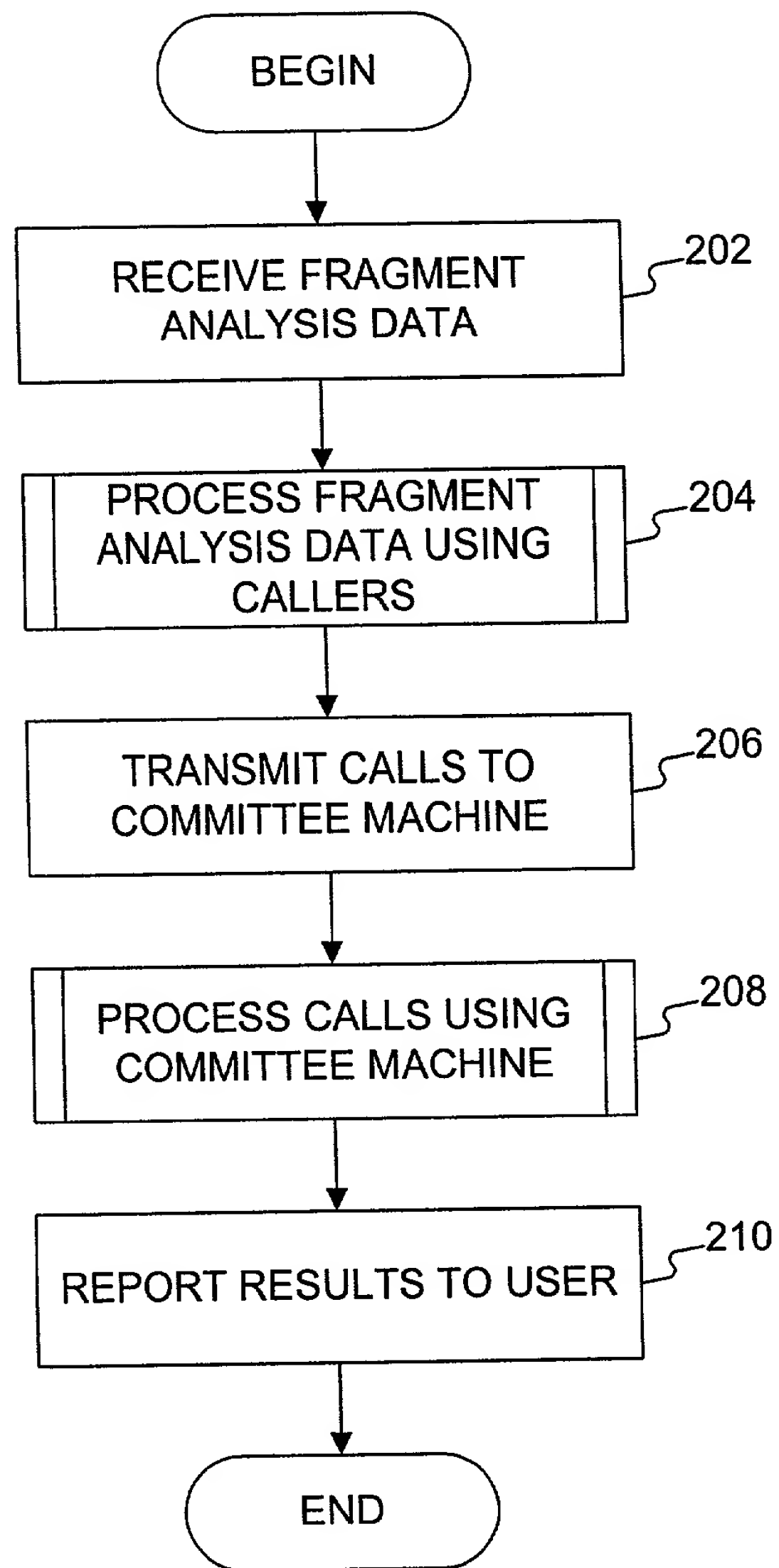
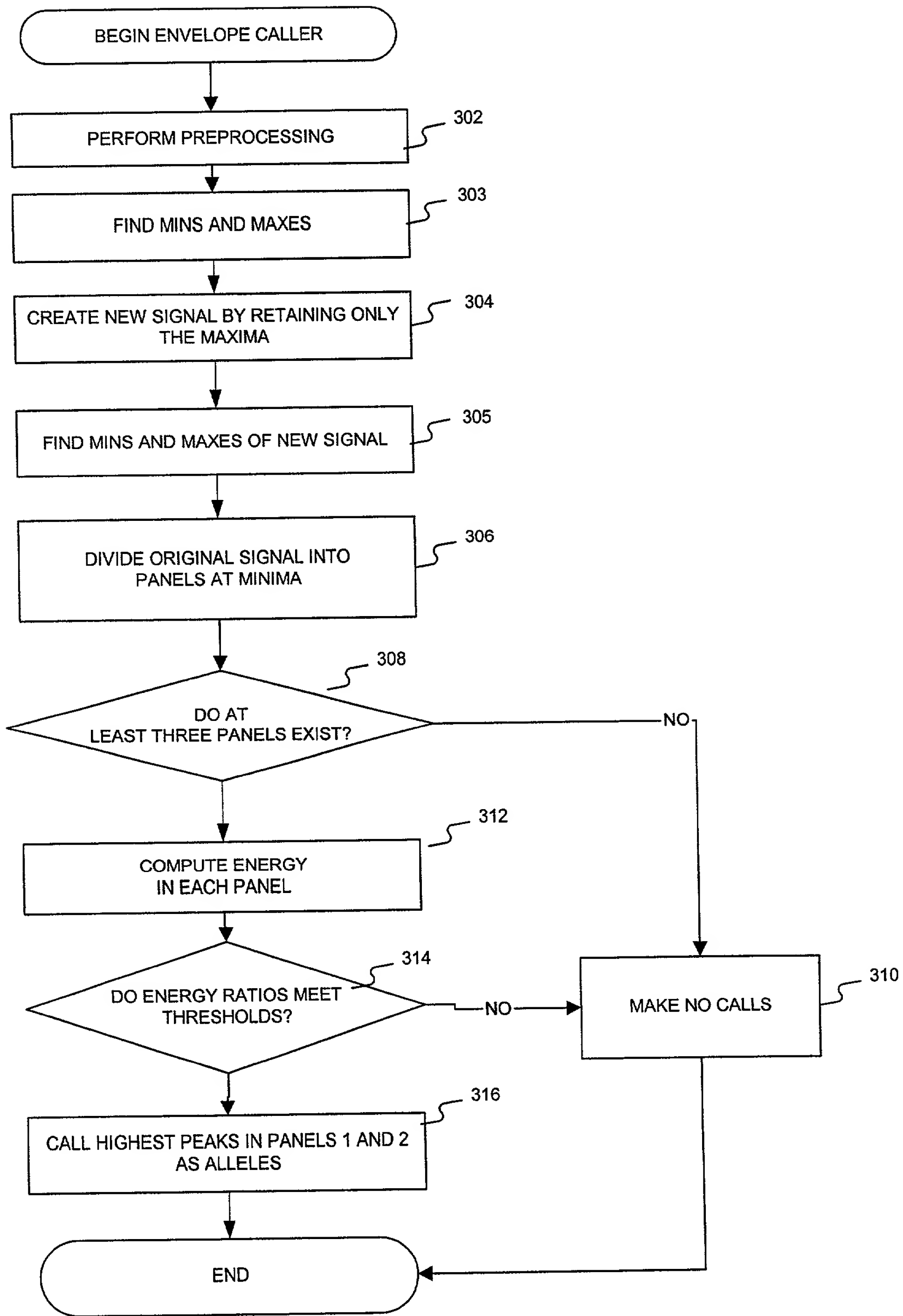


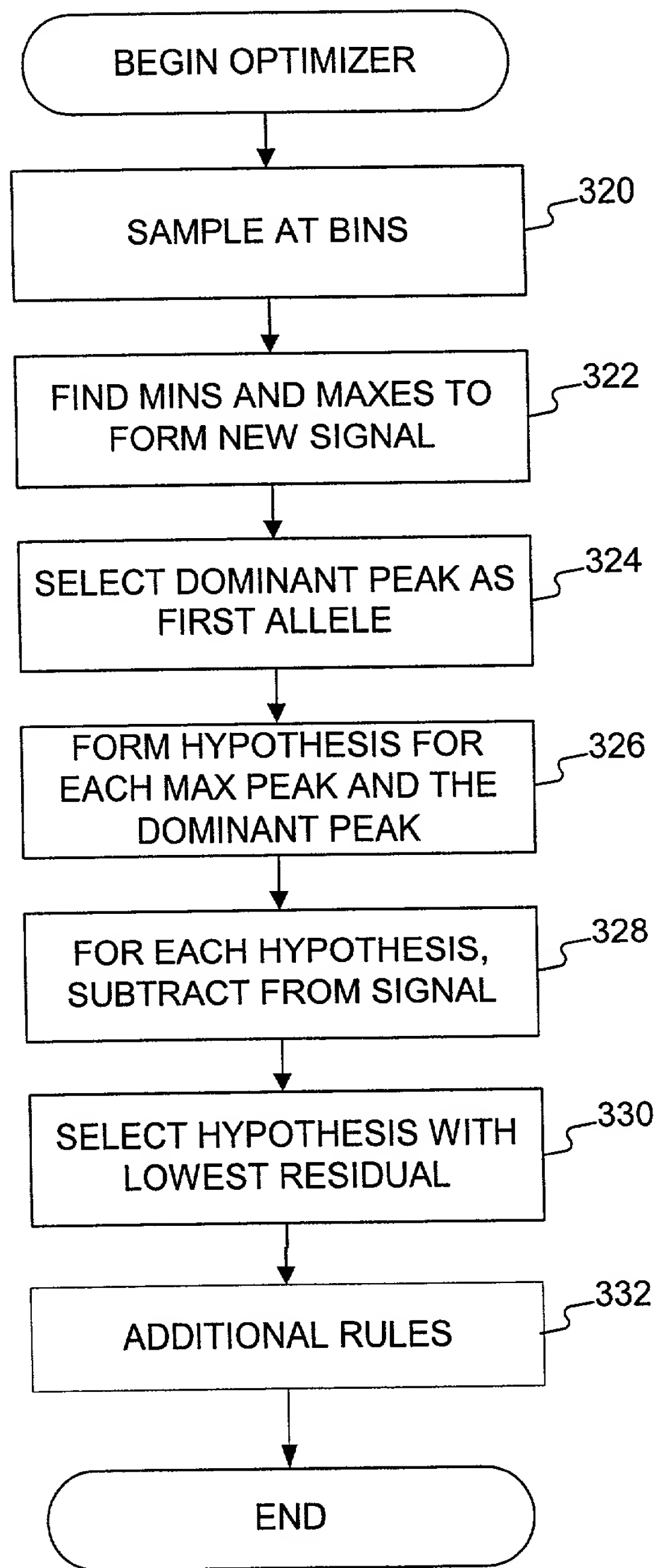
Fig. 1



**Figure 2**

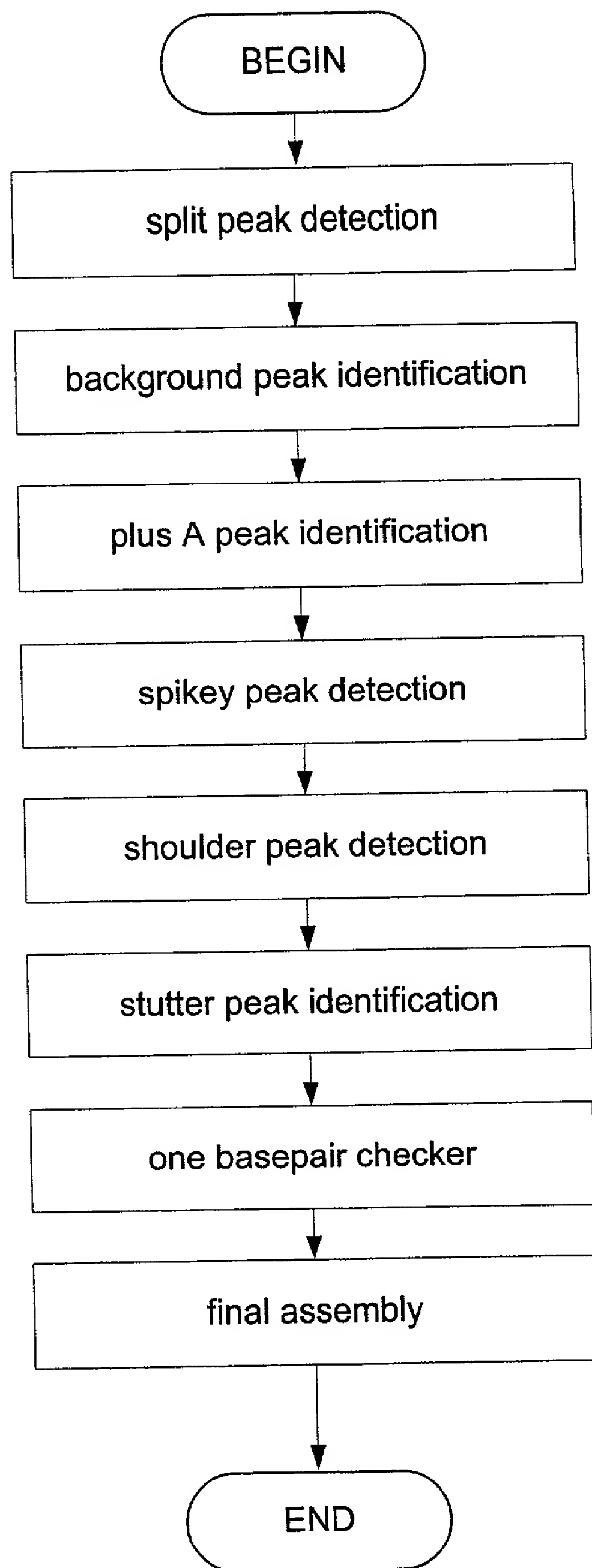


**Figure 3A**

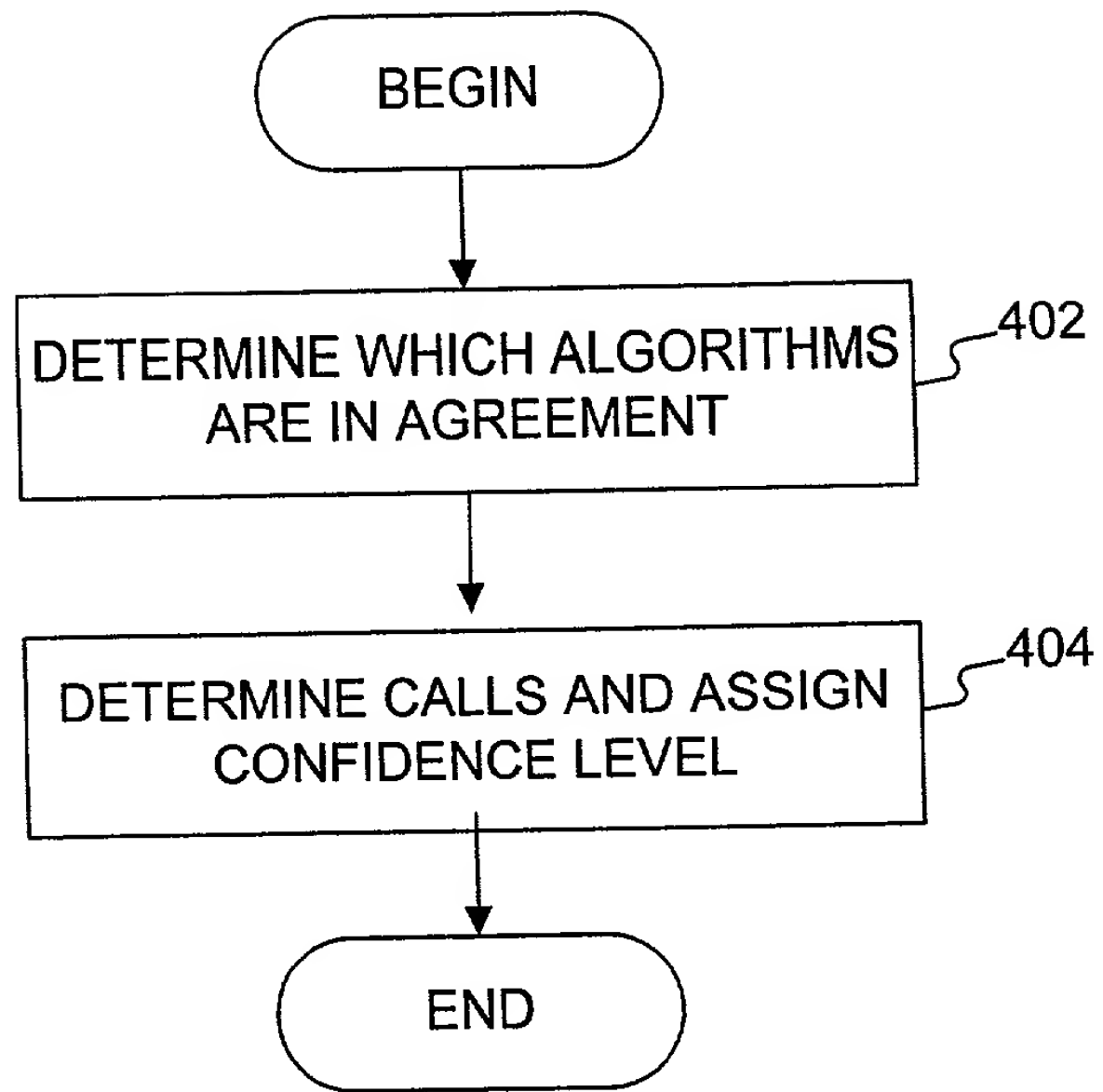


**Figure 3B**

### Figure 3C



**Figure 3D**



**Figure 4**

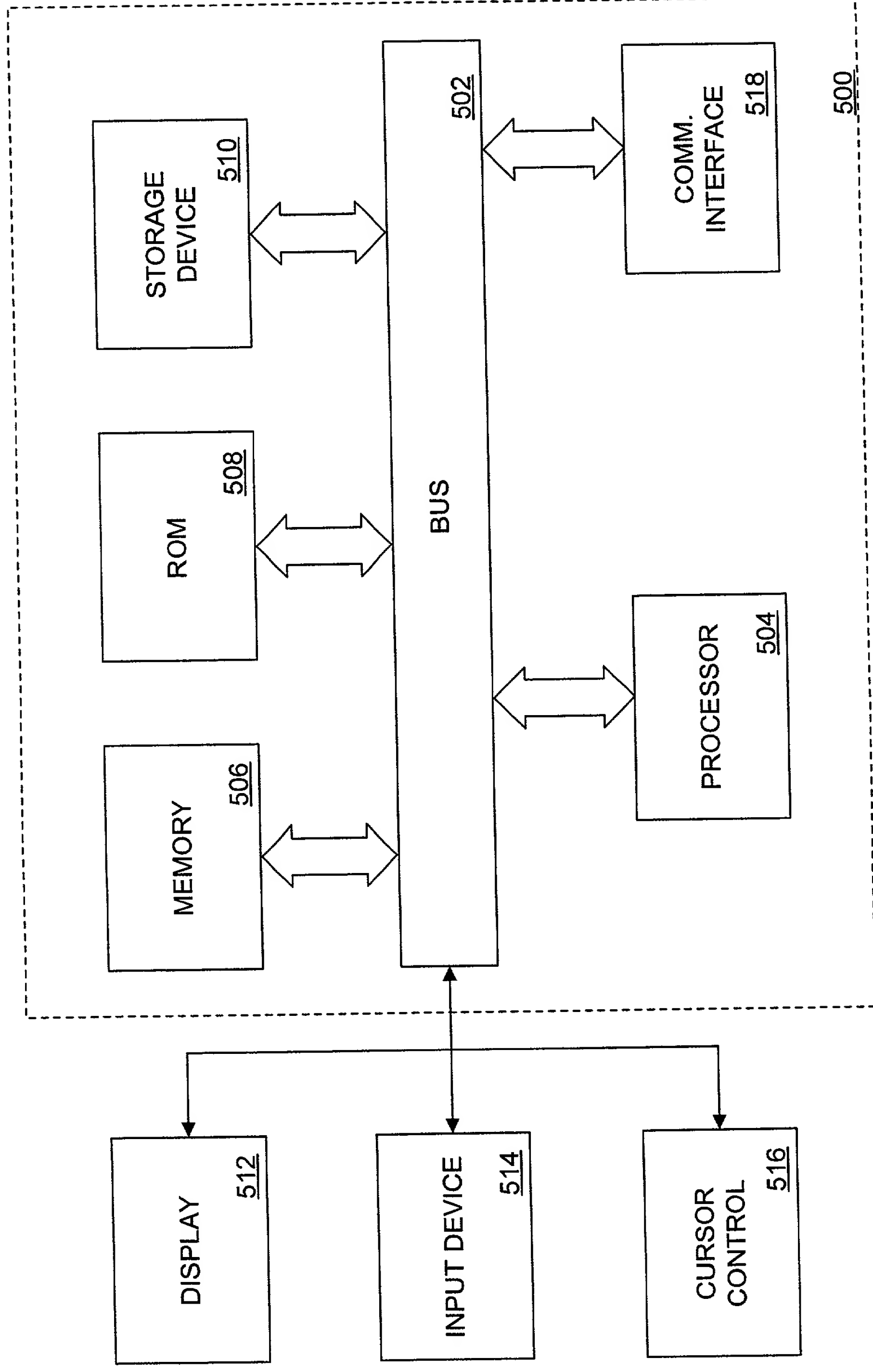


Figure 5



0593061650

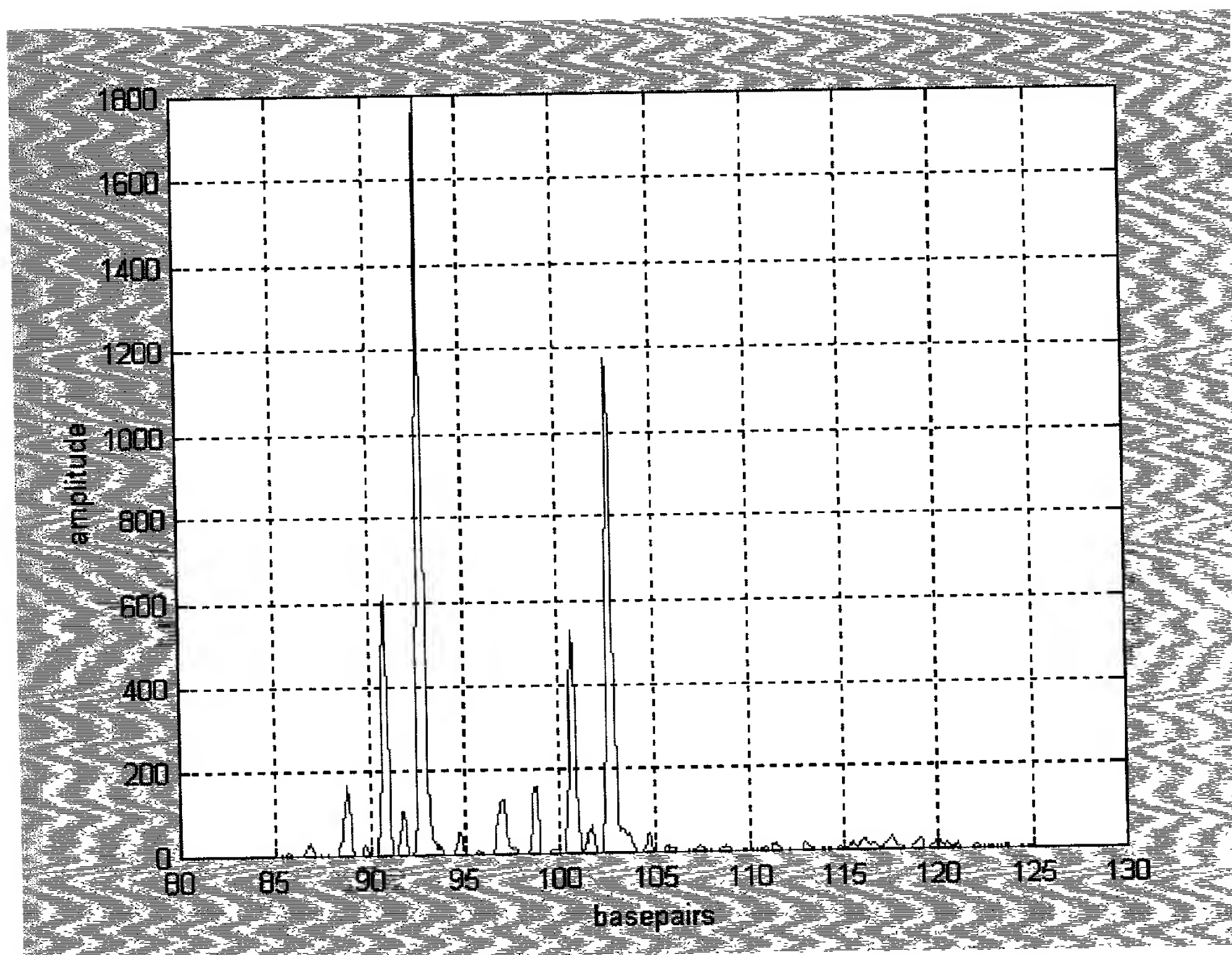


Figure 6

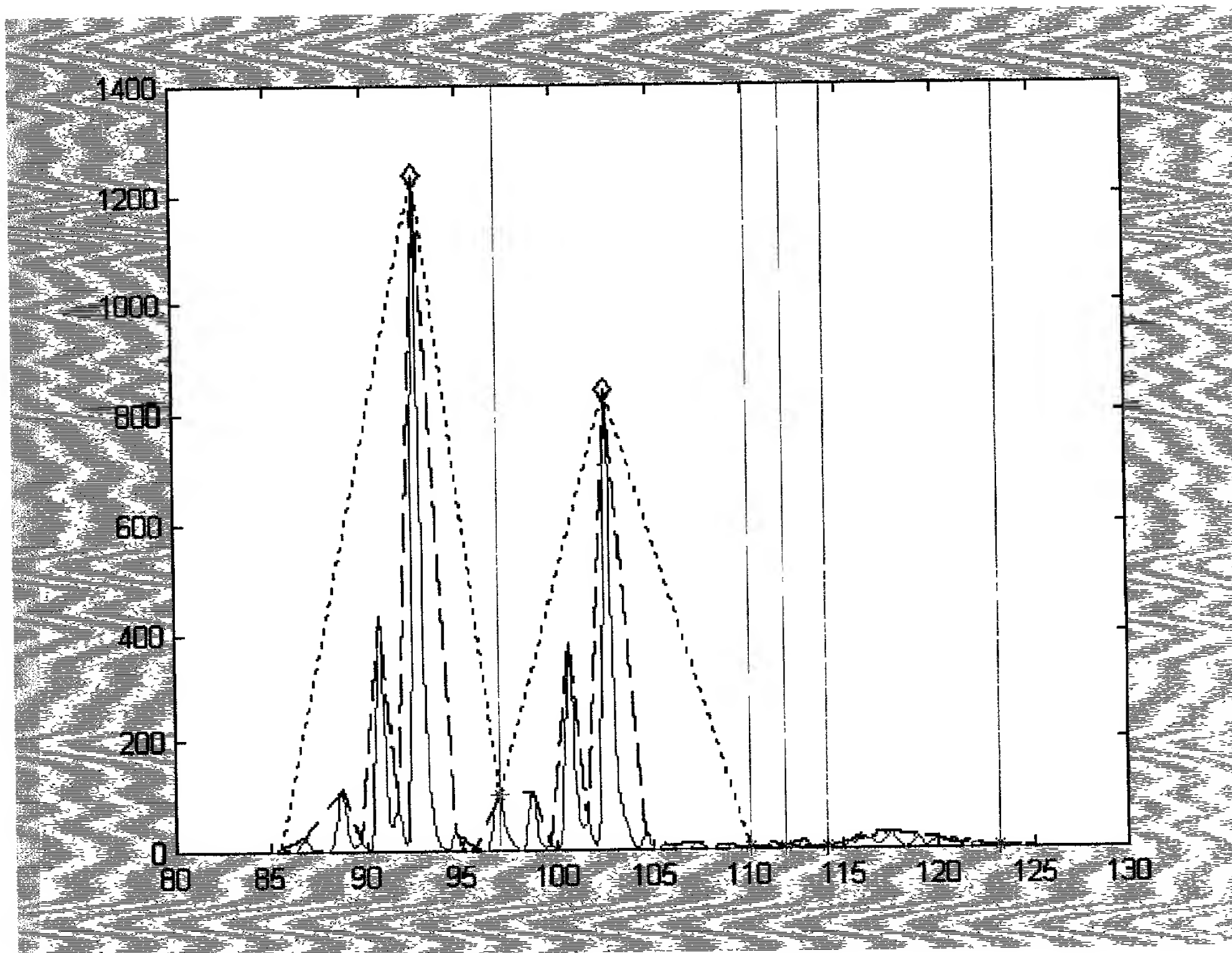


Figure 7

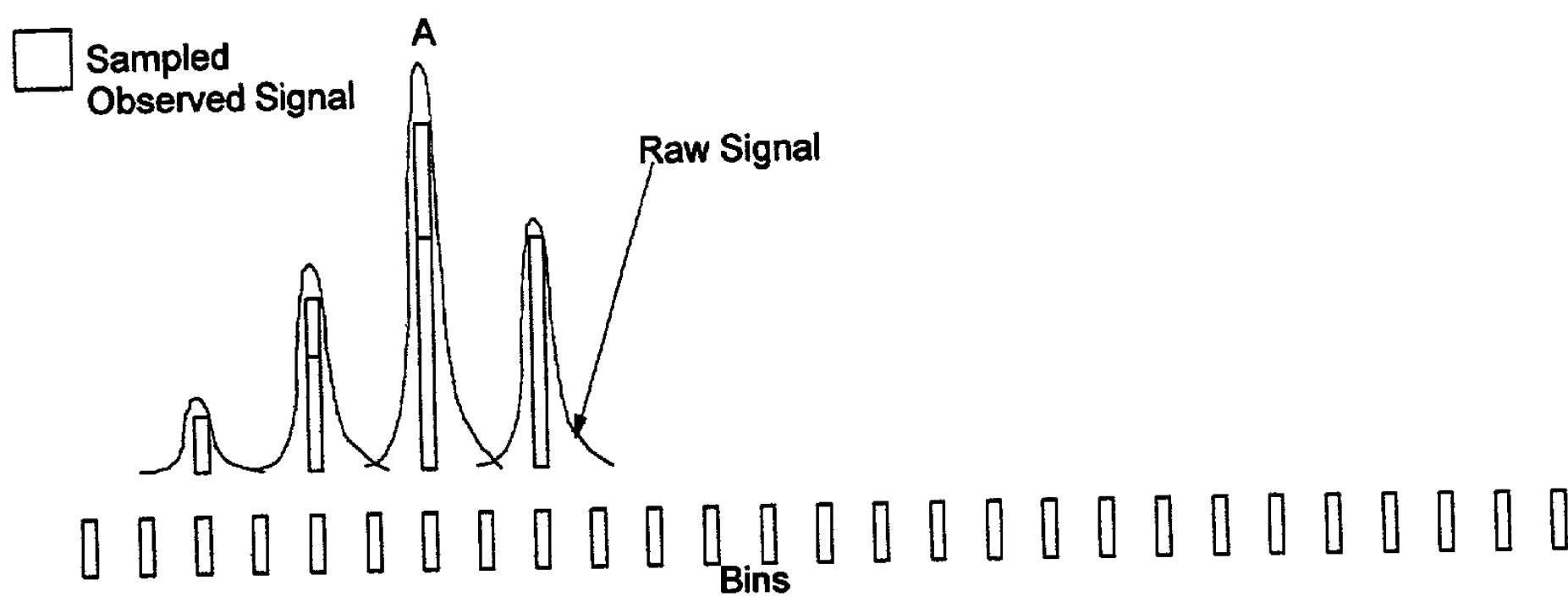


Figure 8

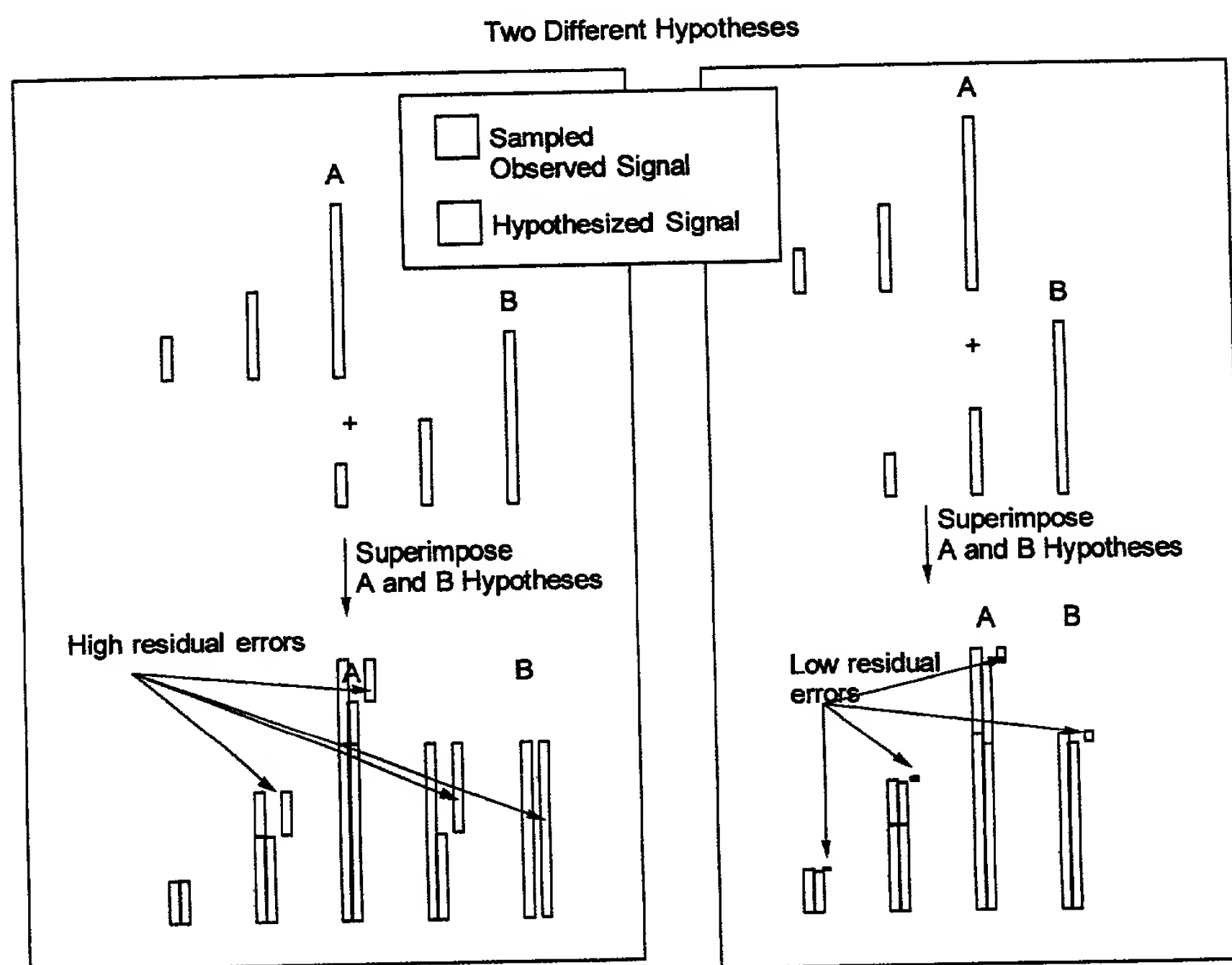


Figure 9

Allele Calling for di-nucleotide marker in linkage mapping  
application Sample Data (3)

Heuristic algorithm ignores the background peaks while calling the correct allele peaks (indicated by red star).

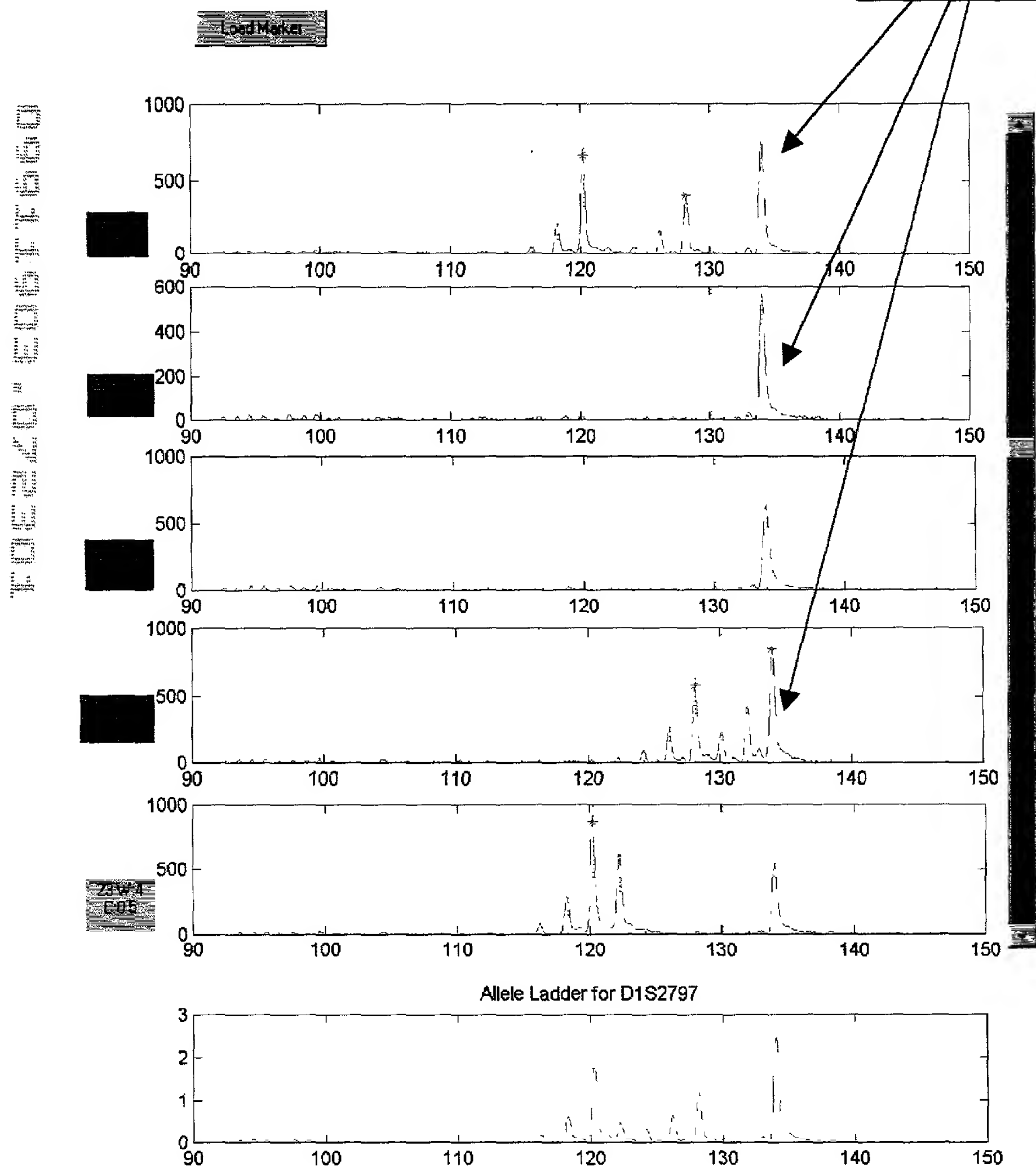
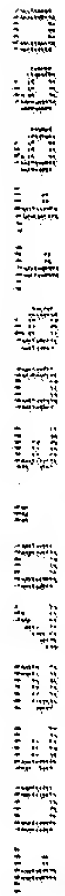


Figure 10

Allele Calling for di-nucleotide marker in linkage mapping application Sample Data (2)

Heuristic algorithm detects the spiky peaks while calling the correct allele peaks (indicated by red star).



### Figure 11

# Allele Calling for di-nucleotide marker in linkage mapping application Sample Data (1)

Heuristic algorithm calls the correct allele peaks (indicated by red star) when serious differential amplification occurs.

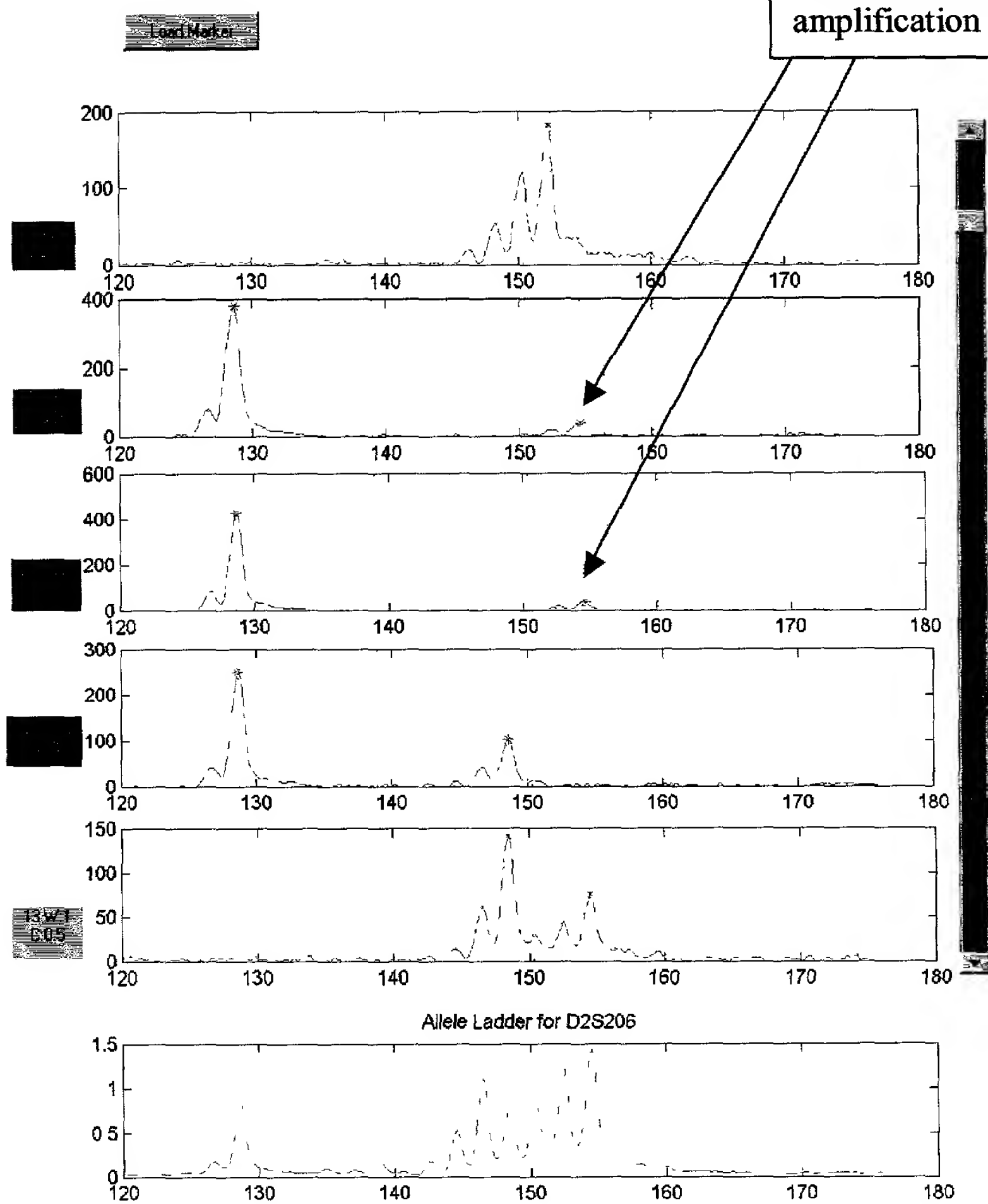


Figure 12

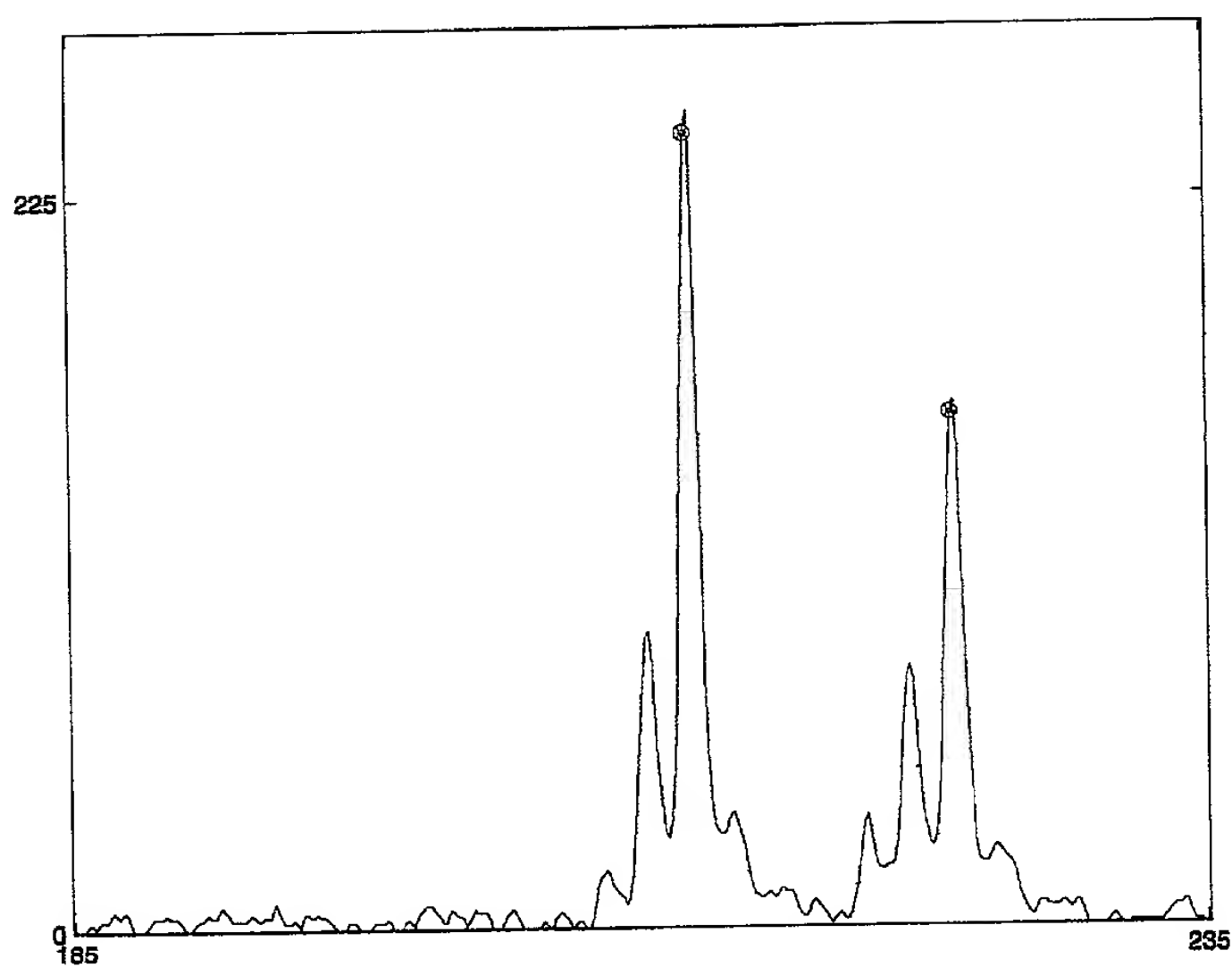


Figure 13: Standard heterozygous allele signature. Circles denote user annotated allele calls. x-axis is in base pairs. y-axis is in A/D counts (voltage intensity)

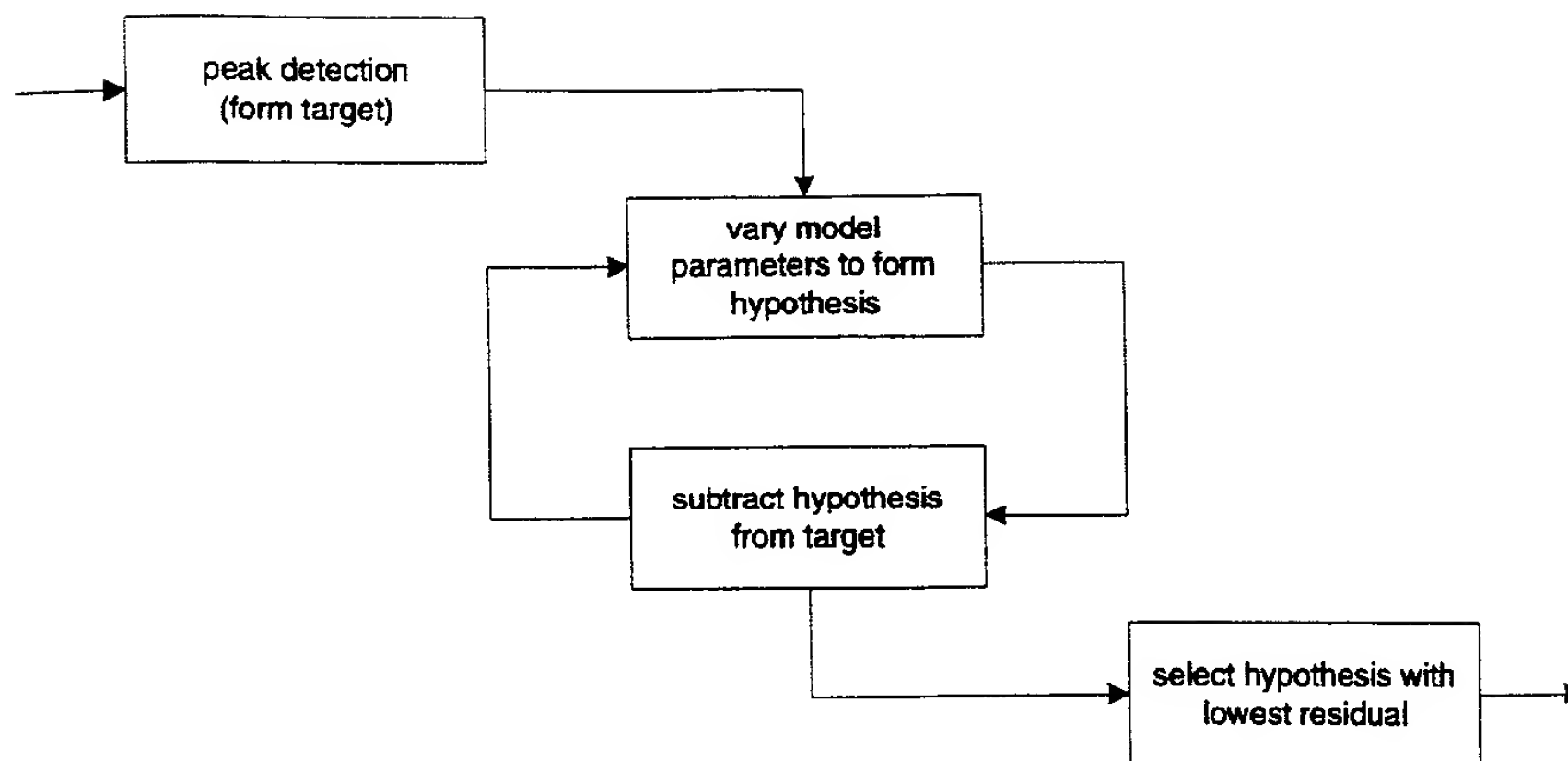


Figure 14: Steps in the allele calling routine. First the signal is simplified via sampling and its peaks are located. This forms the target signal that is to be approximated. The two interconnected boxes indicate the process of varying the parameters and testing how closely the resulting signal matches the sampled version of the original. The set of parameters that yield the closest match contain the allele calls.



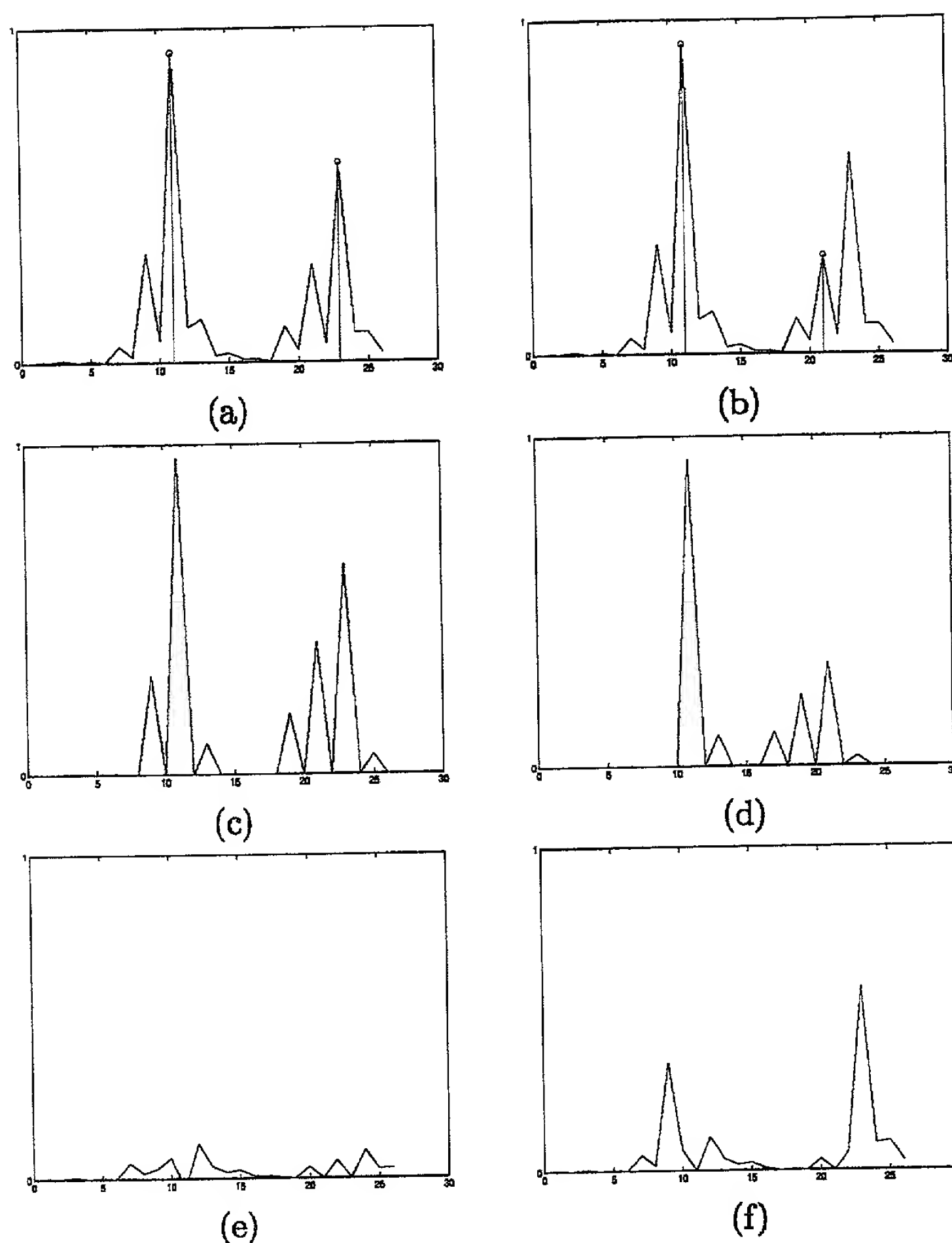


Figure 15: Illustration of hypothesis formation in the optimizer routine. The two columns of Figures represent the optimal solution (left column) and a suboptimal solution (right column). Panel (a) shows the target vector with the two red lines showing the location of the candidate peaks. Panel (c) shows the hypothesis formed using different values of stutter and  $^+A$ . Panel (e) shows the residual error resulting from subtraction of the signal in panel (c) from the signal in panel (a) (sum squared error = 0.0355). Panels (b,d,f) show the same process for a slightly different allele hypothesis. This is a poor hypothesis and the residual is rather significant (SSE = 0.4715). The x-axis is somewhat meaningless at this point since it gets mapped back to base-pair indices after the winning hypothesis is chosen.

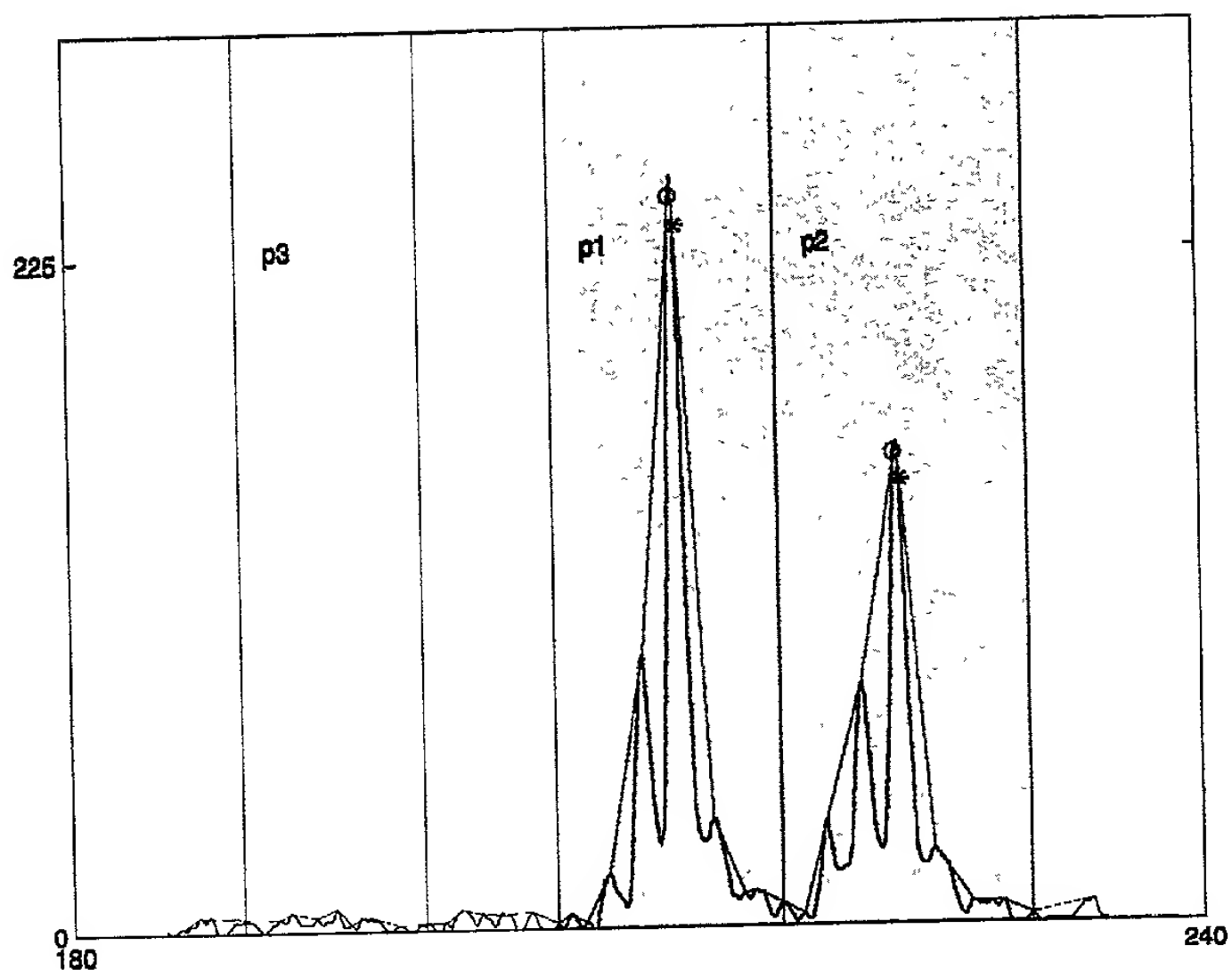
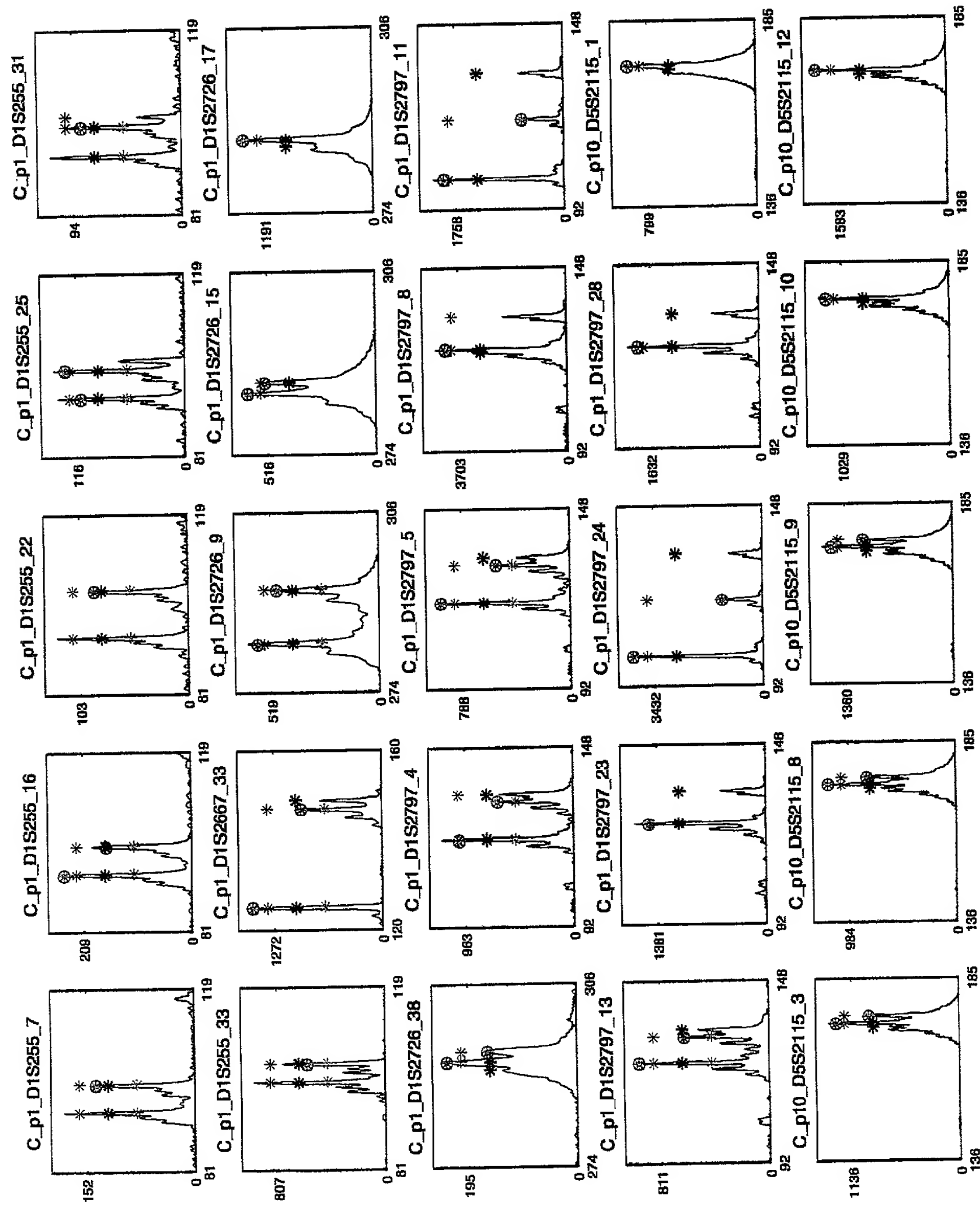


Figure 16: Division of heterozygous signal into panels by the Envelope Caller algorithm. The panels are ranked according to signal energy and the three of interest are labeled p1, p2 and p3 with the two panels containing strong allele signatures being shaded in blue. Circles denote user annotated allele calls. x-axis is in base pairs. y-axis is in A/D counts (voltage intensity)



**Figure 17: Example of how reporting could be accomplished. These are examples where consensus was not reached and show data that is difficult to interpret.**

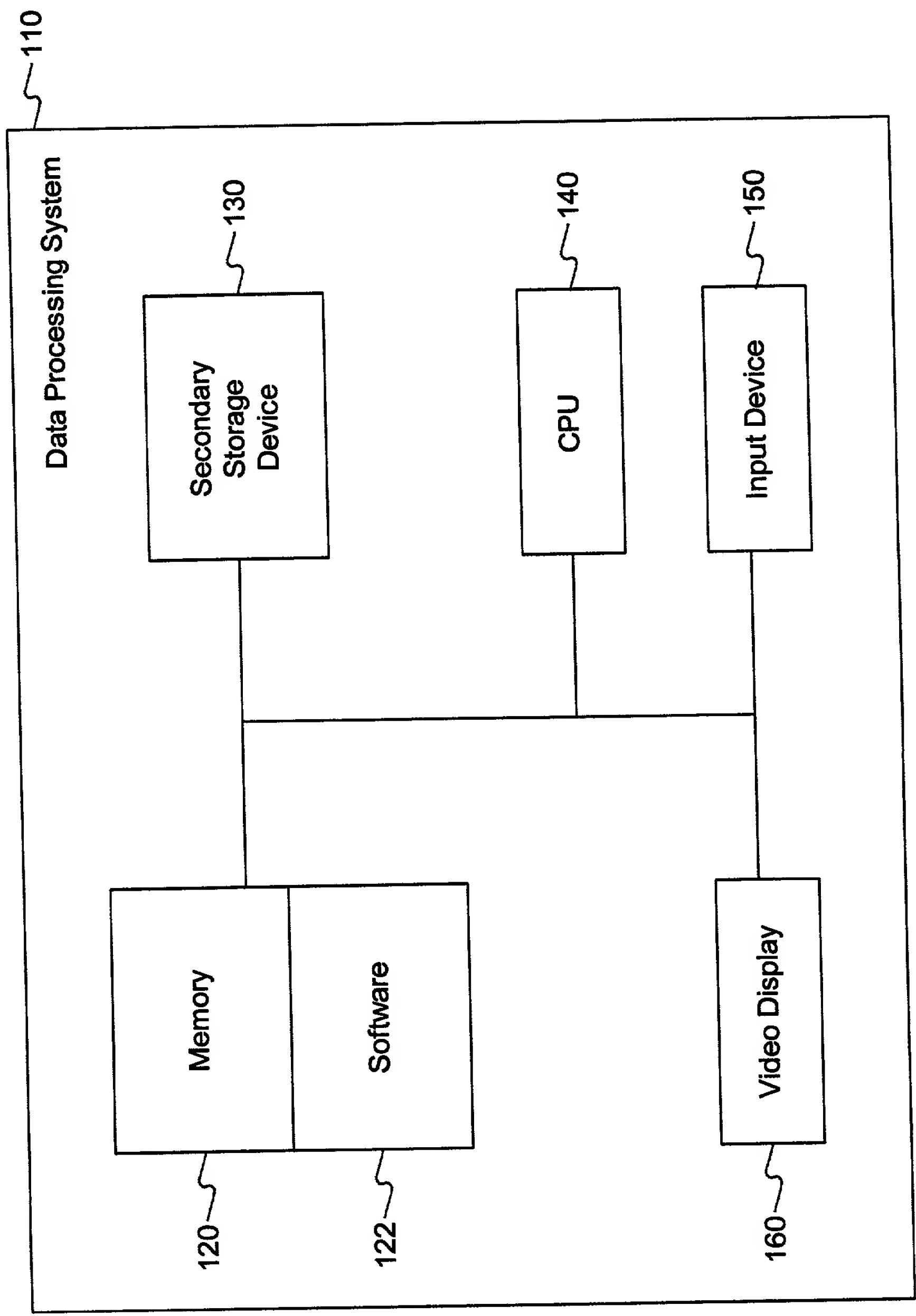


Figure 18

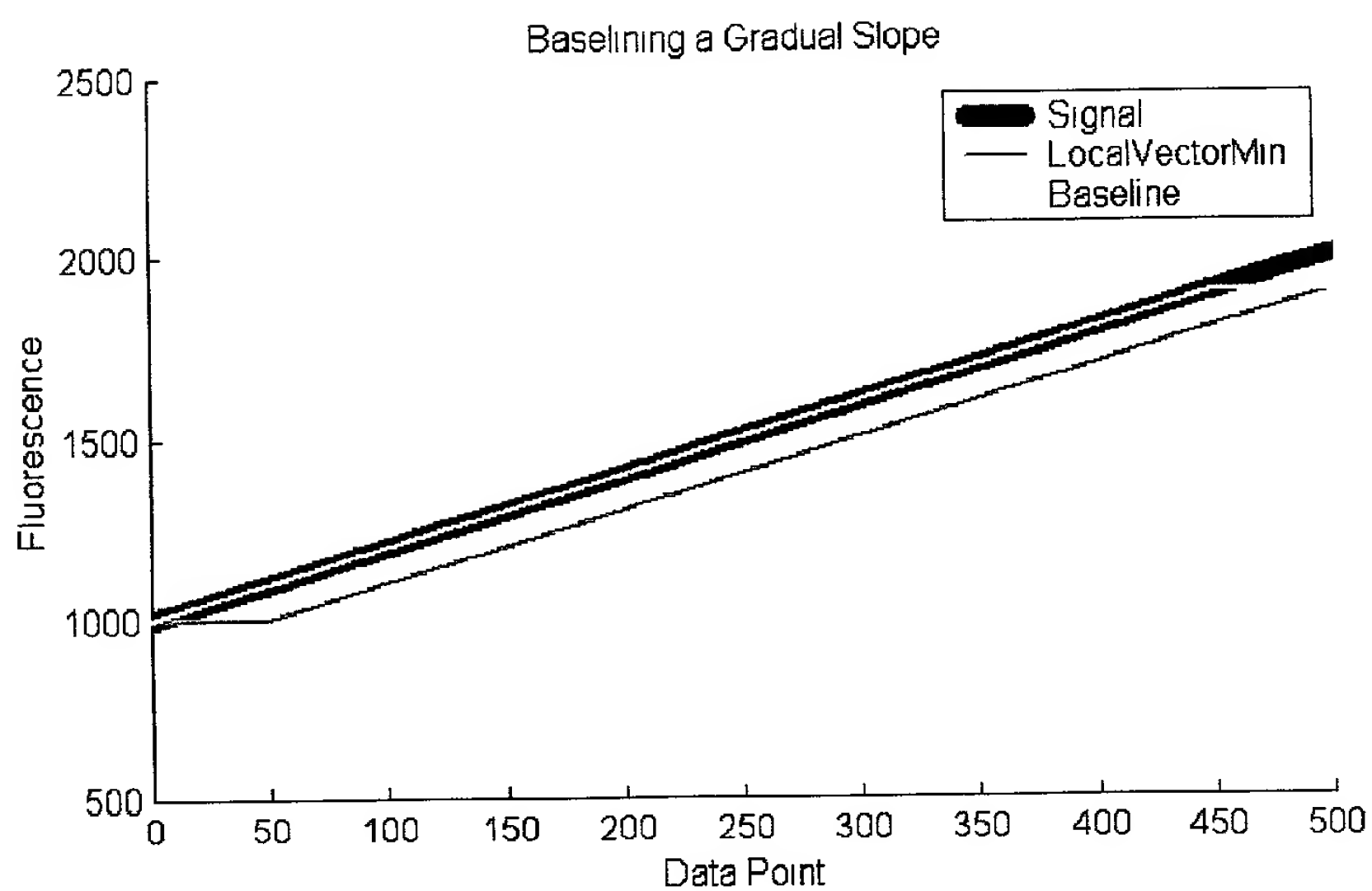


Figure 19

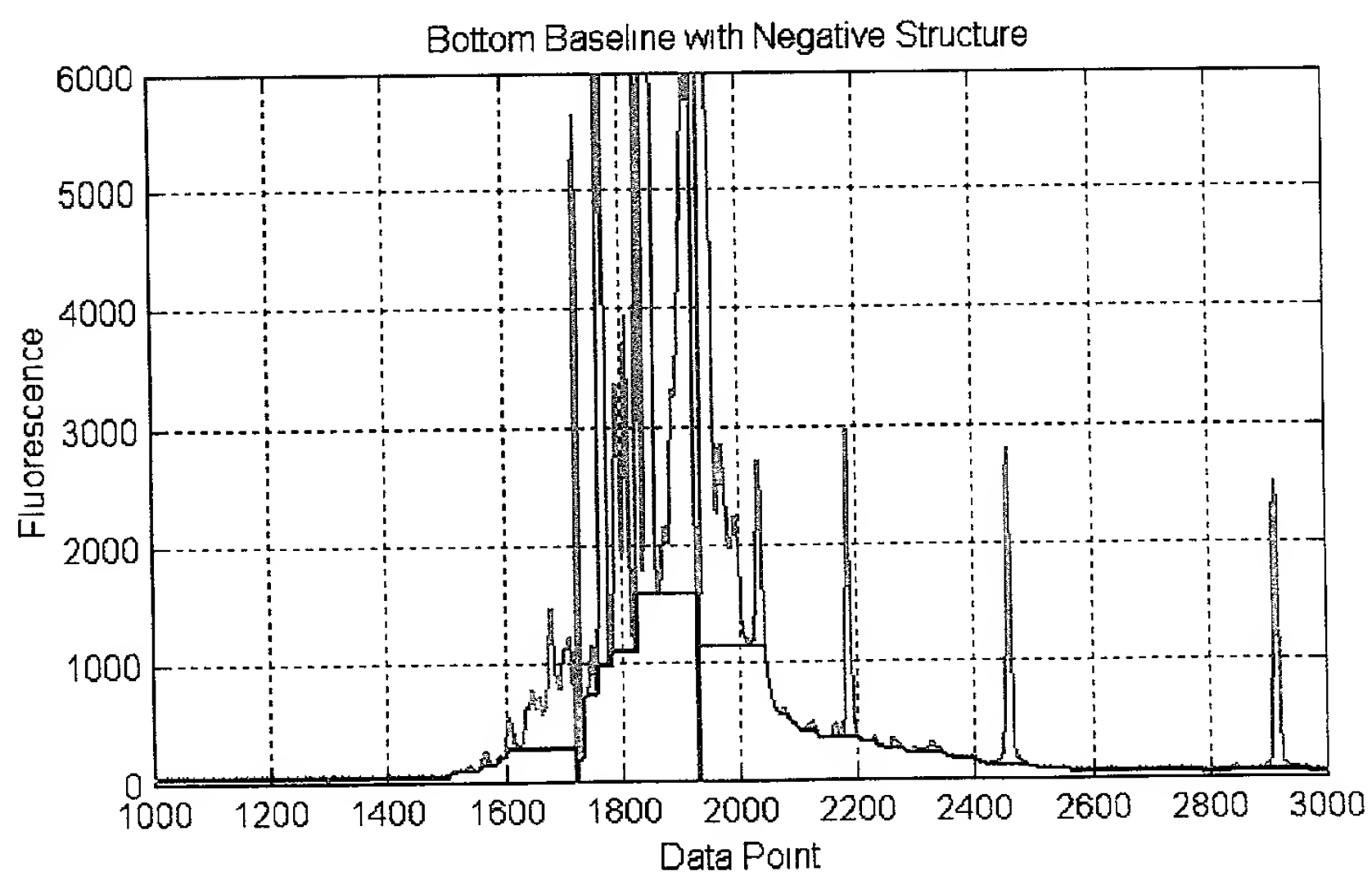


Figure 20: The bottom baseline is shown in blue. Note the two downward spikes, around data points 1700 and 1900 in the primer peak.

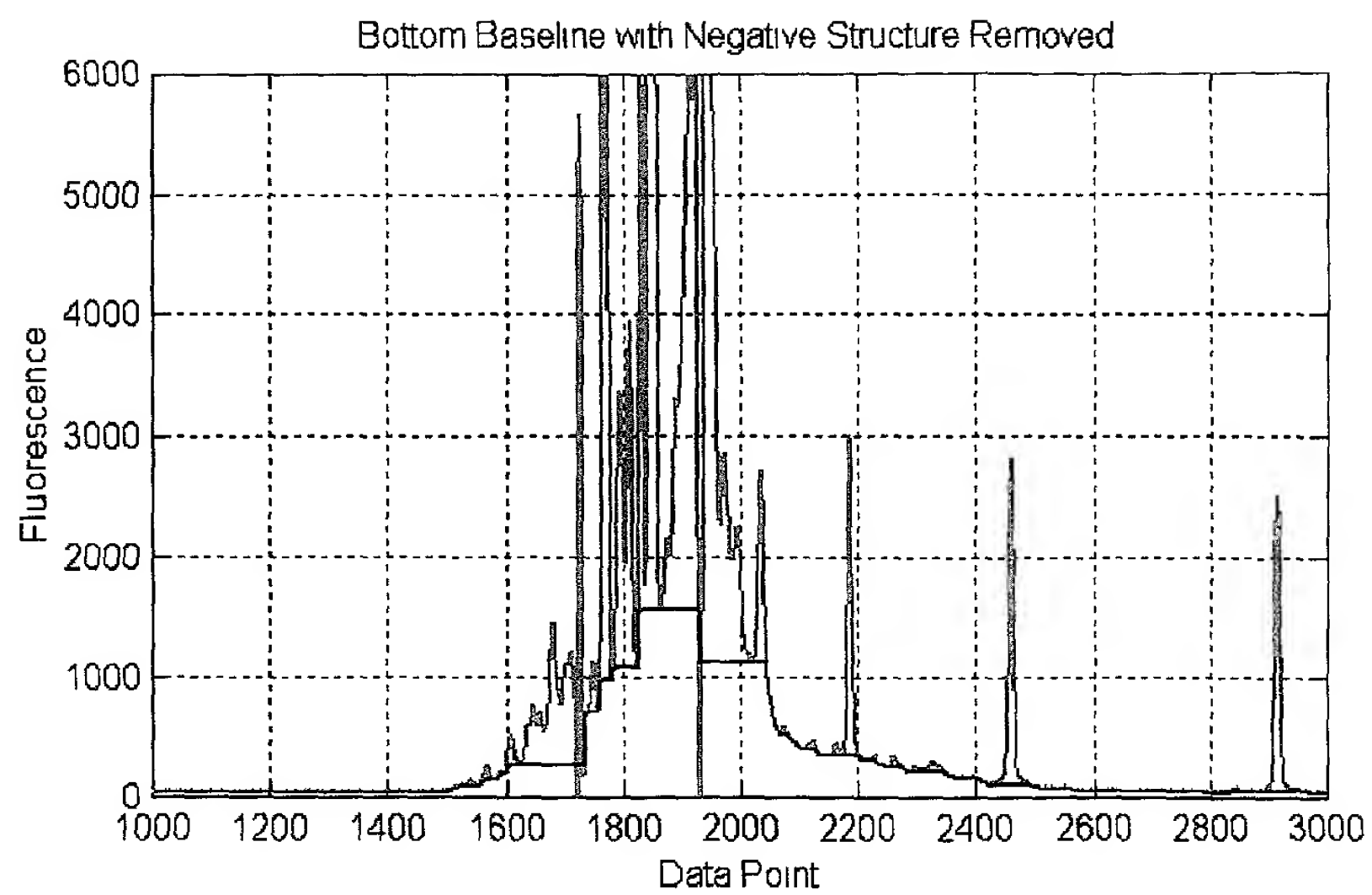


Figure 21: Bottom baseline after eliminating the negative spikes.

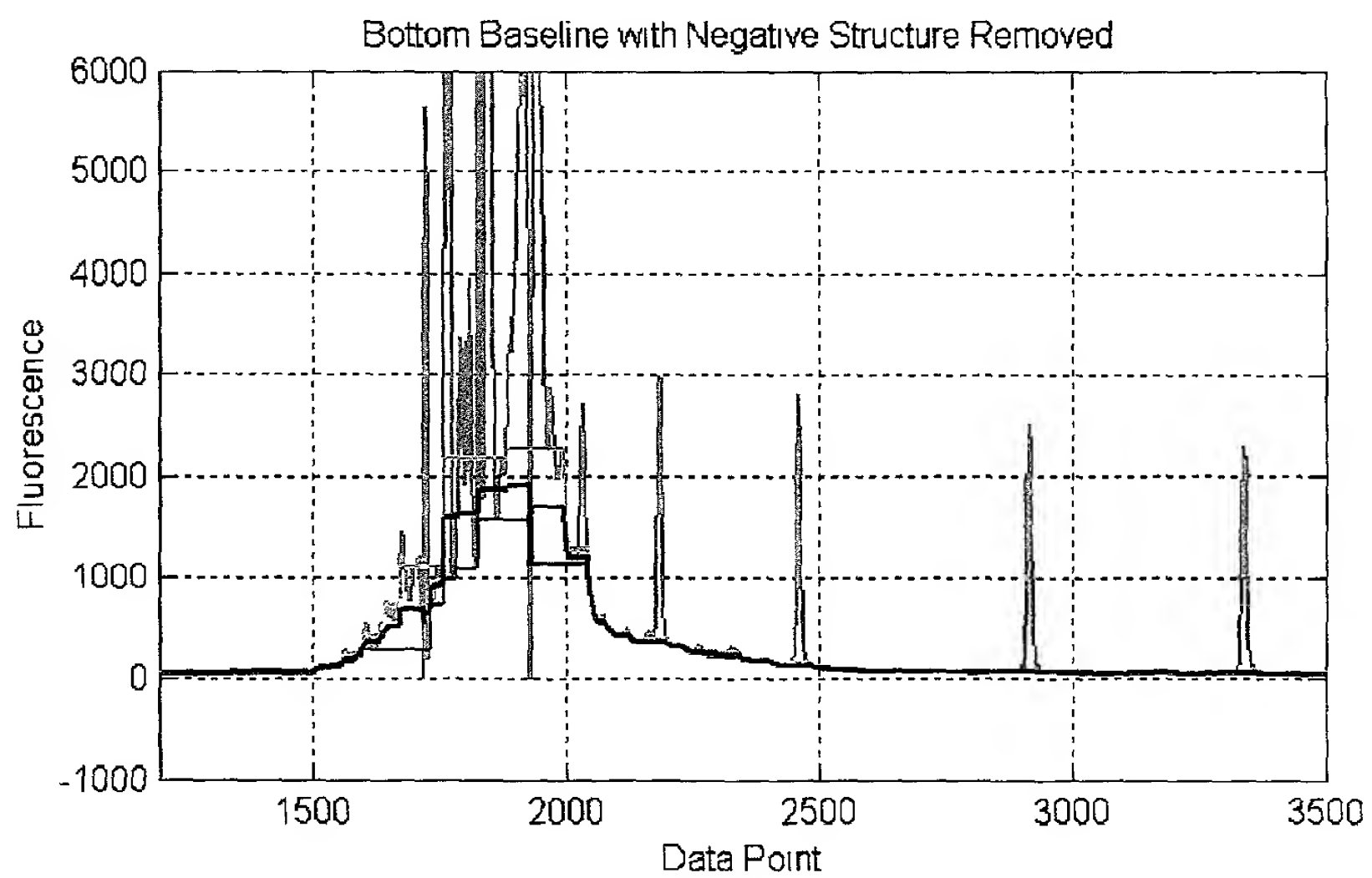


Figure 22: Baseline by averaging the top and bottom.

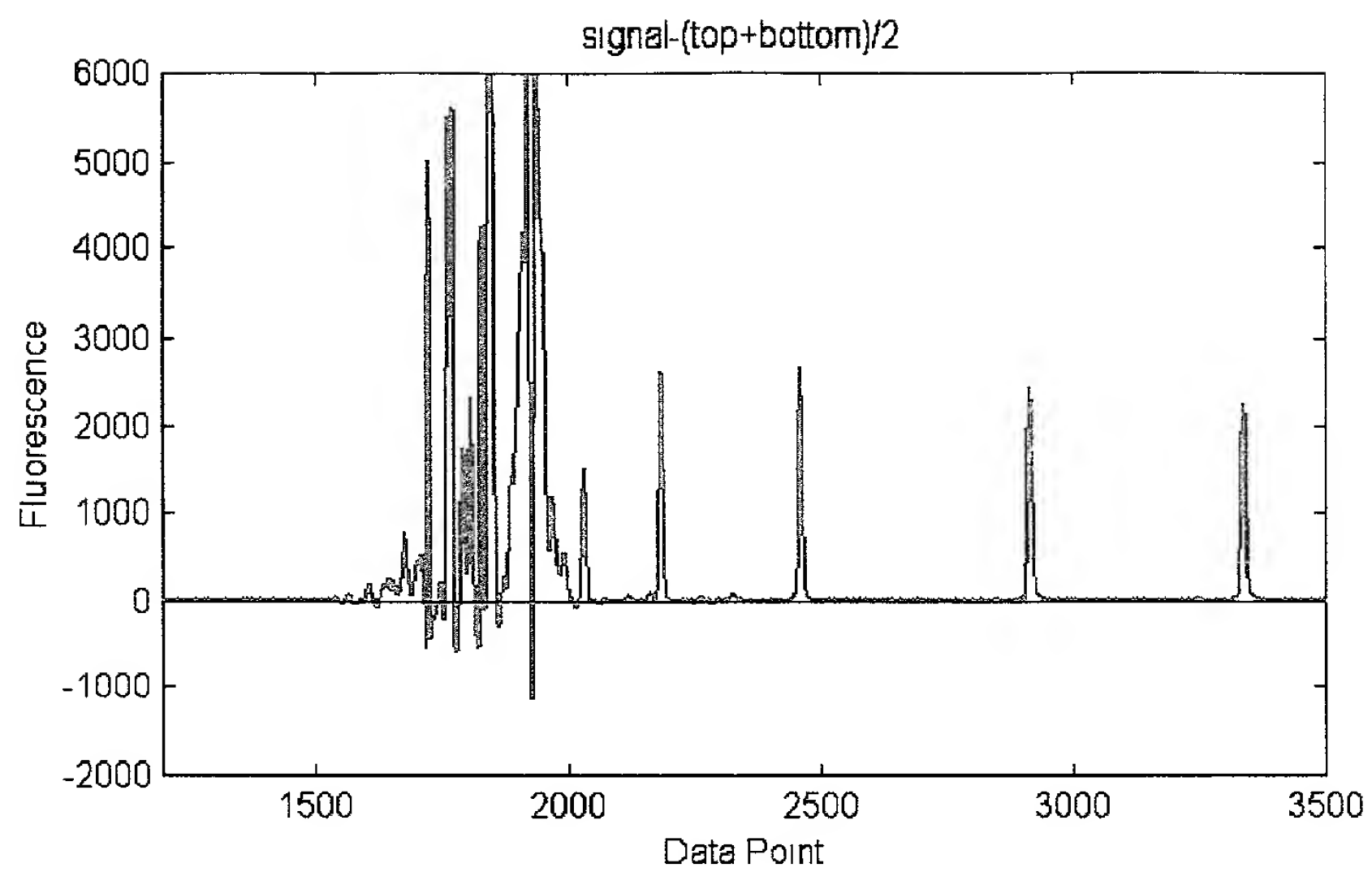


Figure 23: The baselined signal. Note negative values.





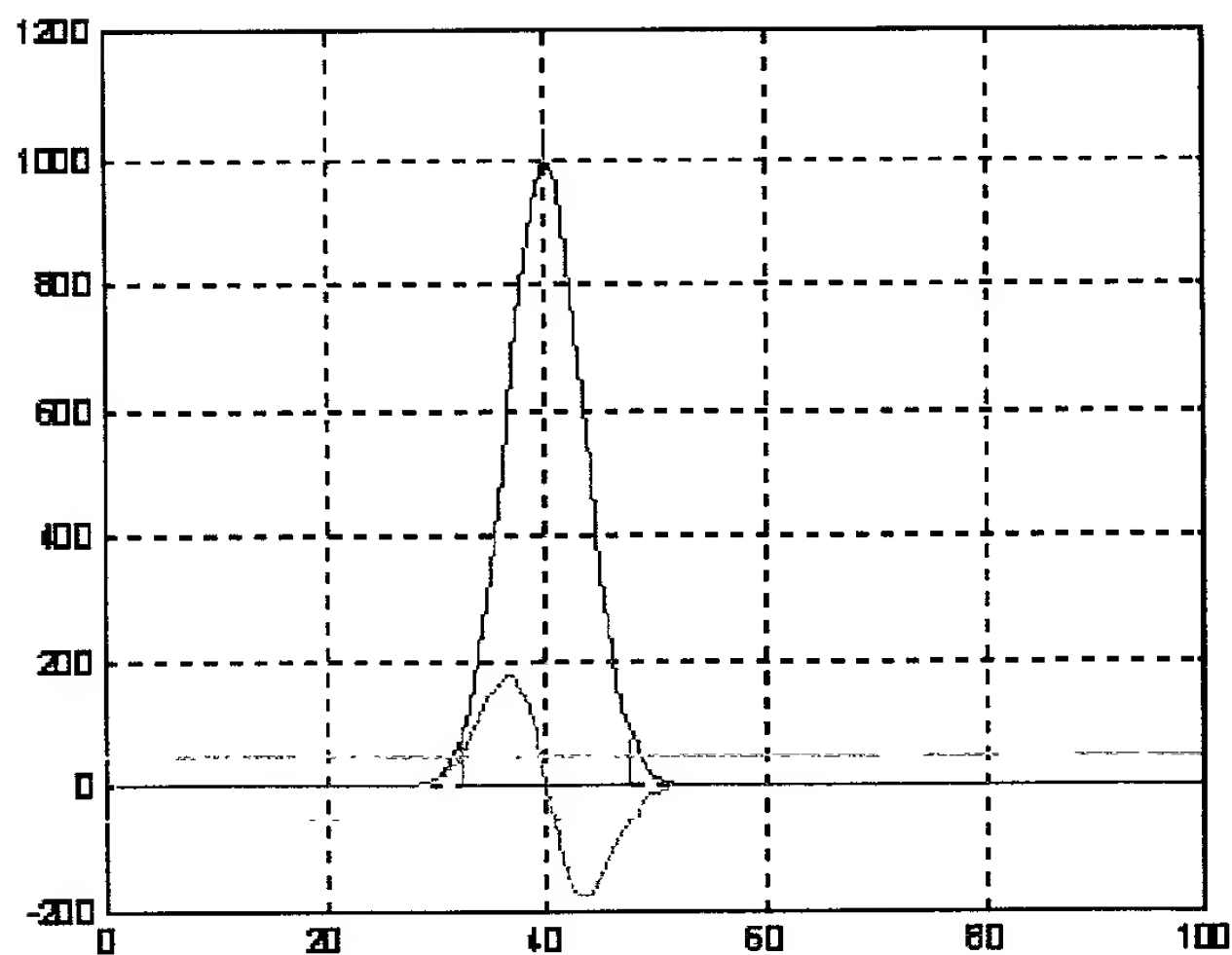


Figure 26: Set symmetric parameters if the peak is symmetric.

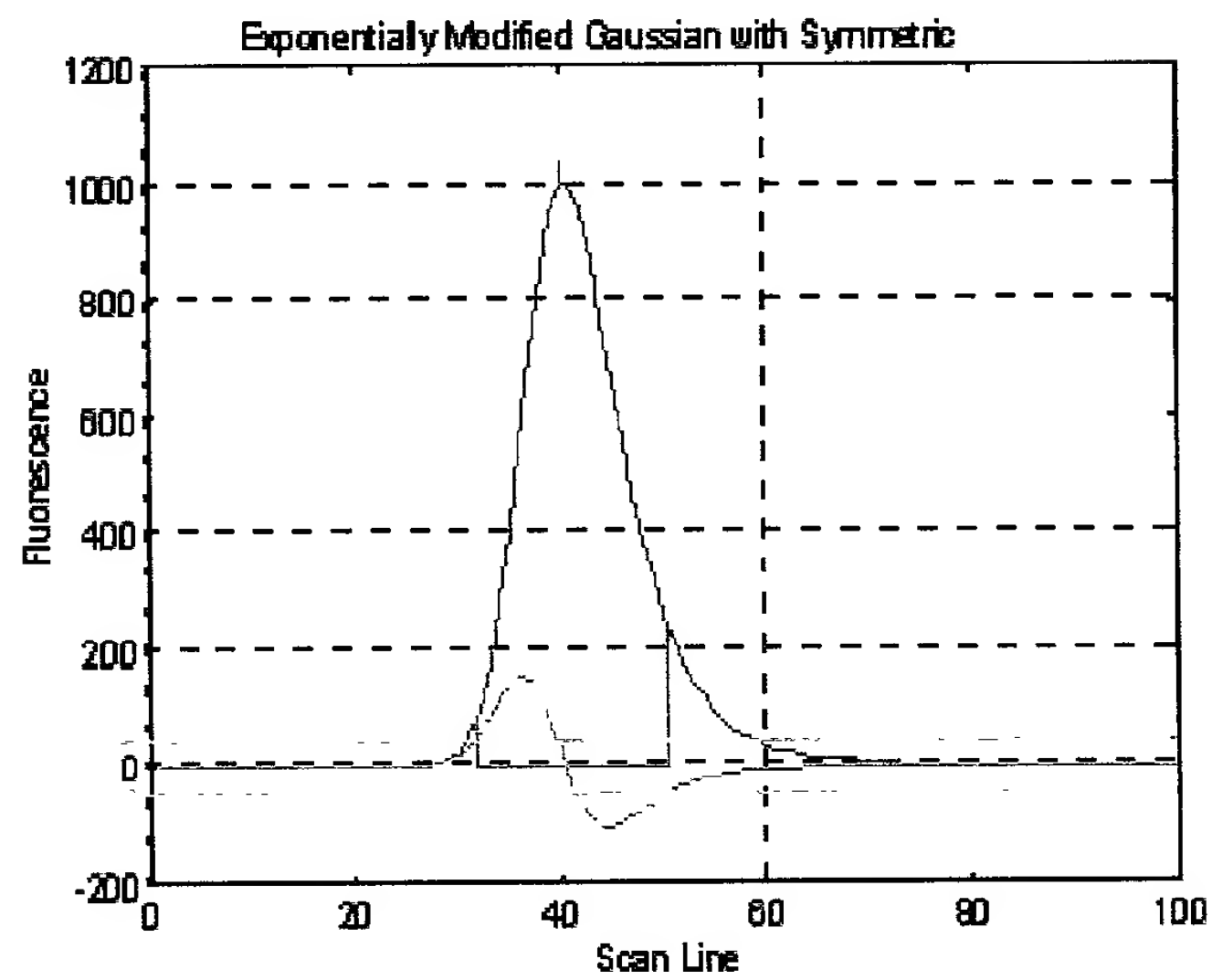


Figure 27: Do not set symmetric parameters if the peak is not symmetric.

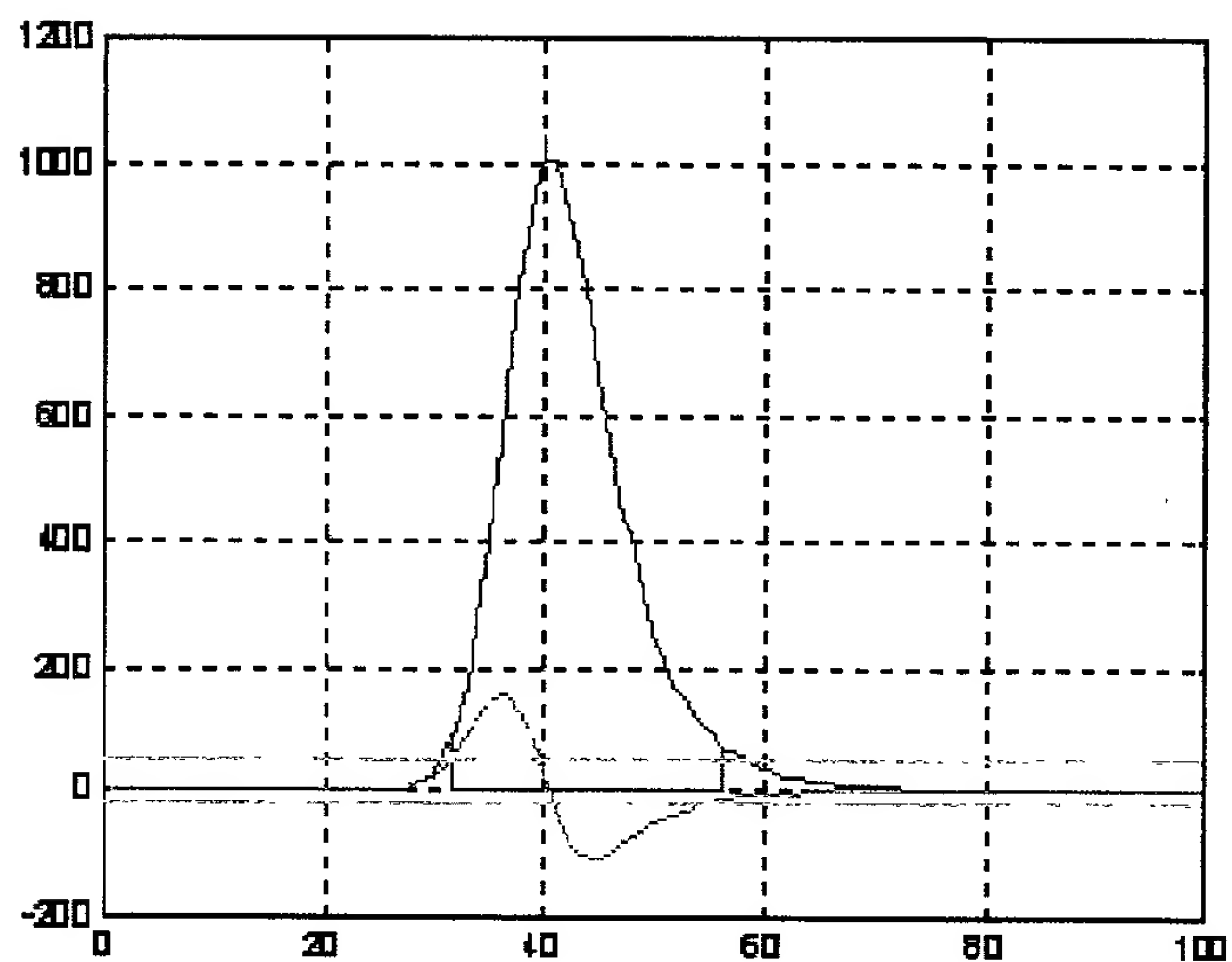


Figure 28: Set asymmetric parameters if the peak is not symmetric.

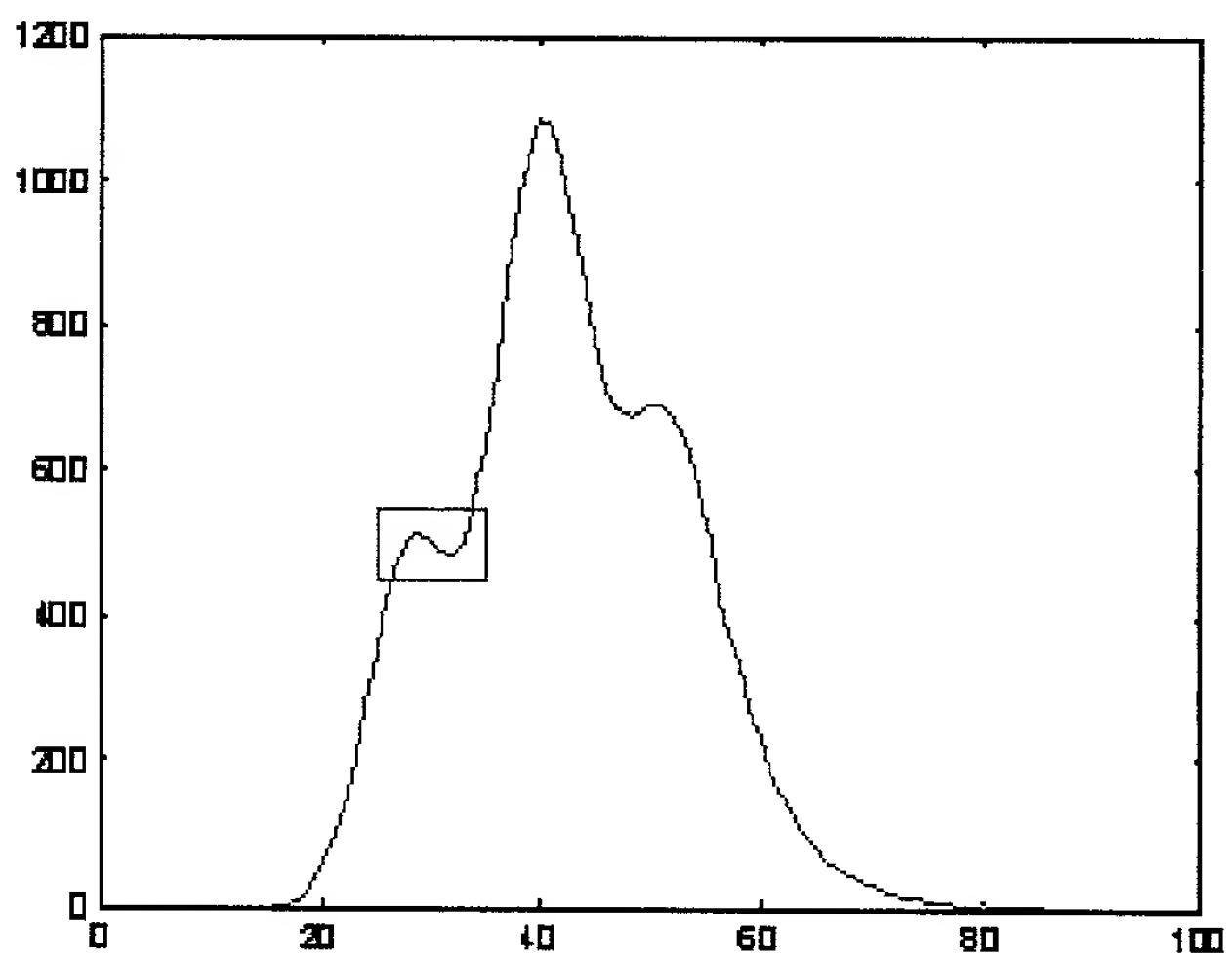


Figure 29: Calculating the first derivative at  $x = 30$ .

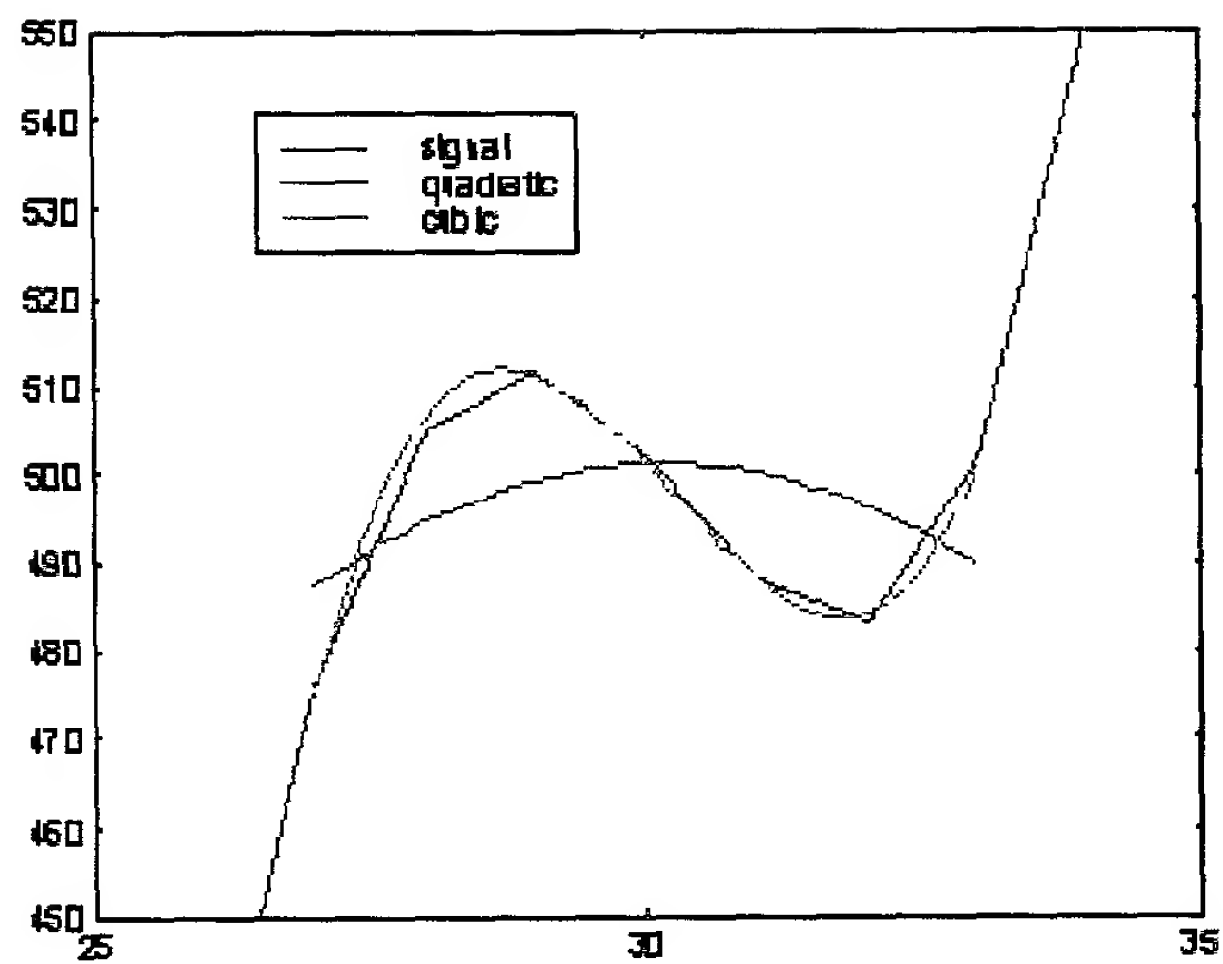


Figure 30: Calculating the first derivative by fitting polynomials.

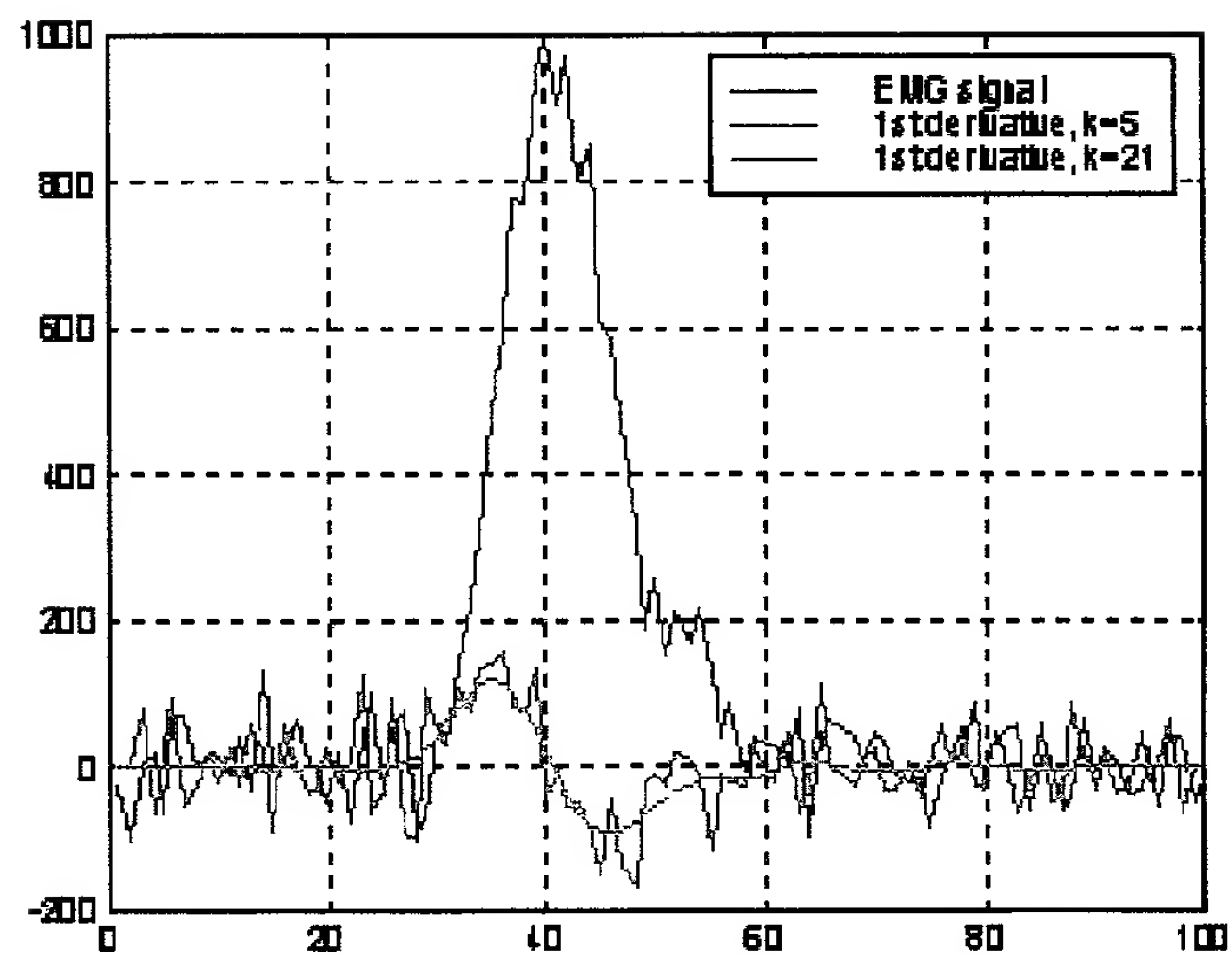


Figure 31: Using  $k$  to smooth the derivative.

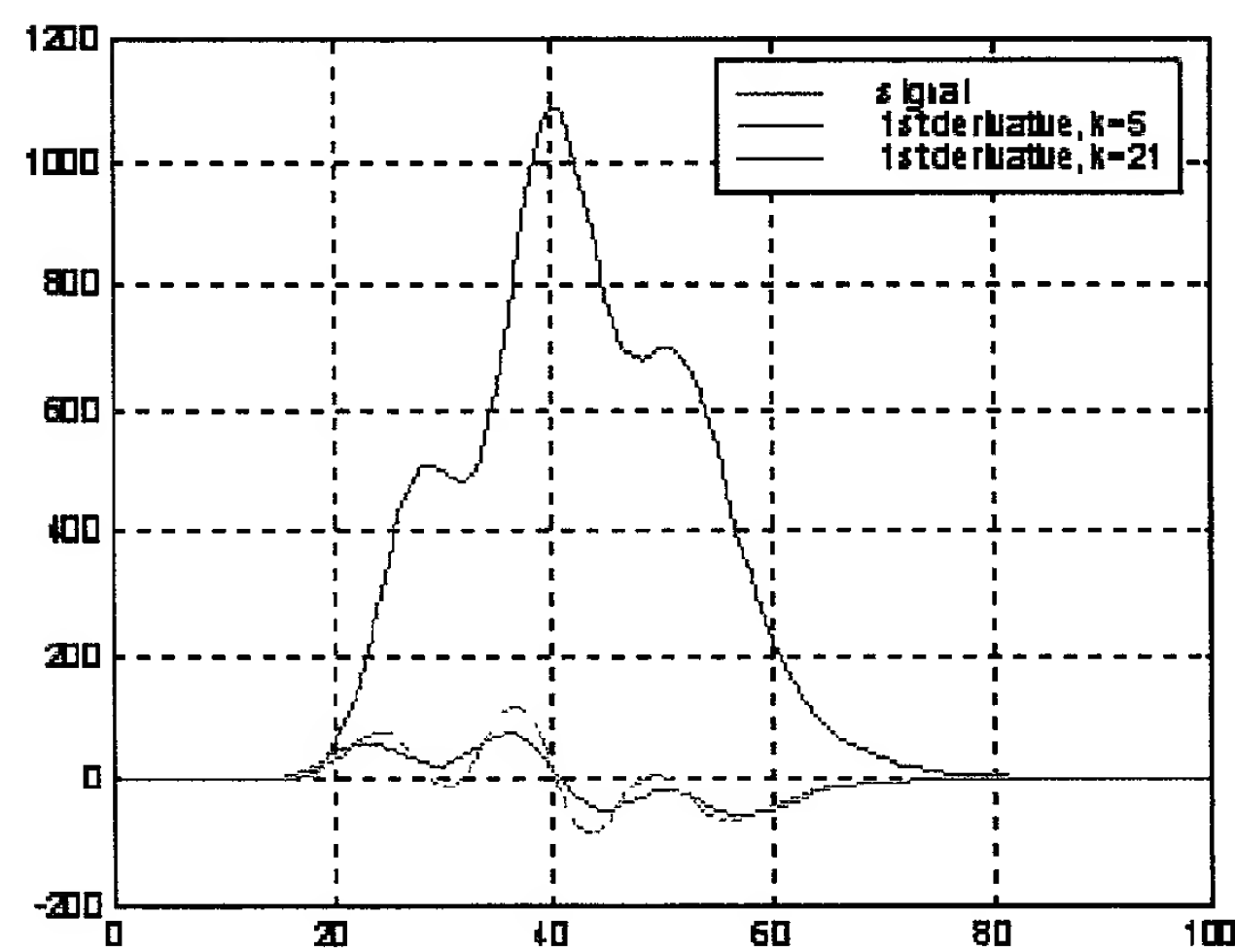


Figure 32: Three peaks for  $k = 5$ , one peak for  $k = 21$ .

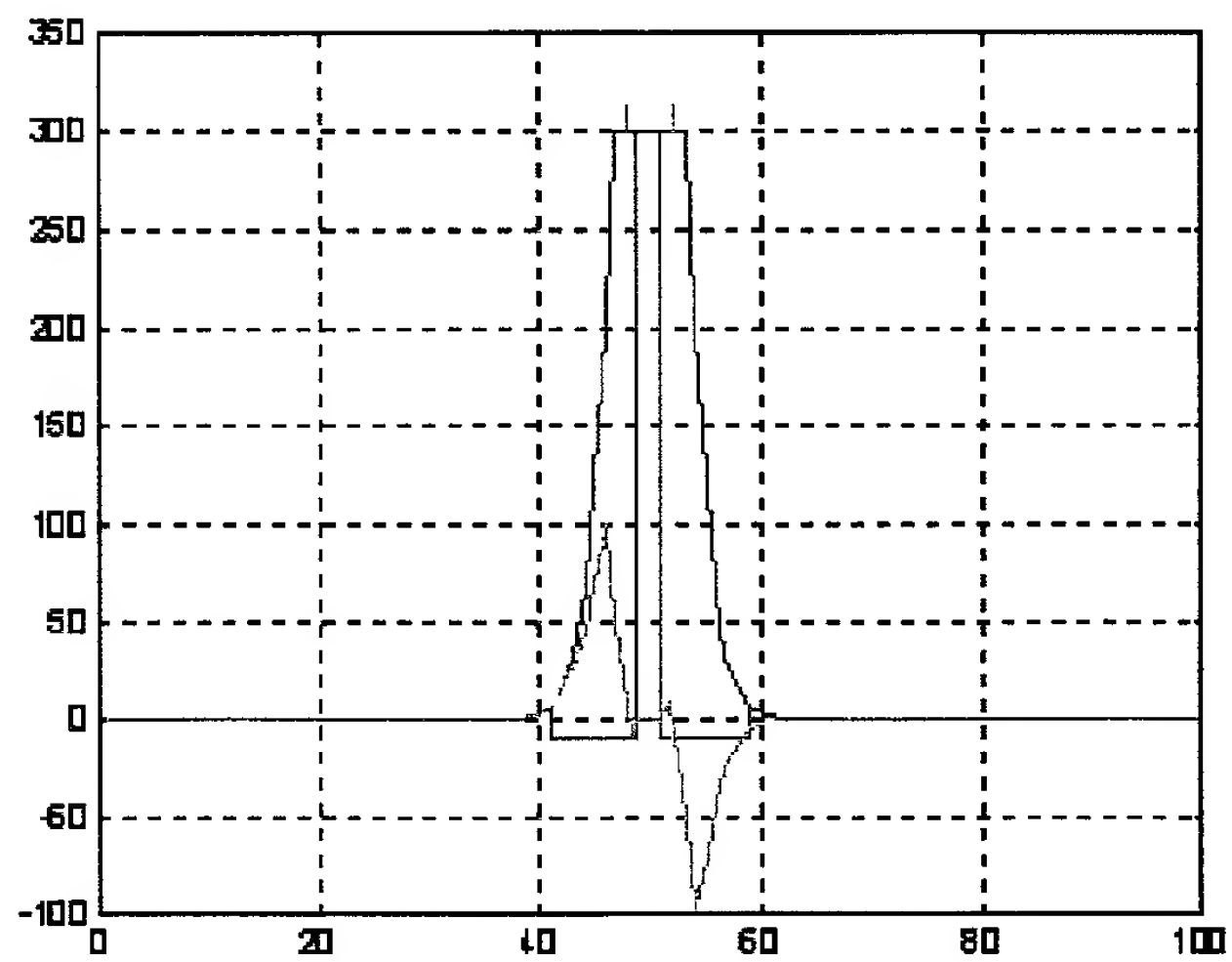


Figure 33: Twin towers detected (where only one exists.)

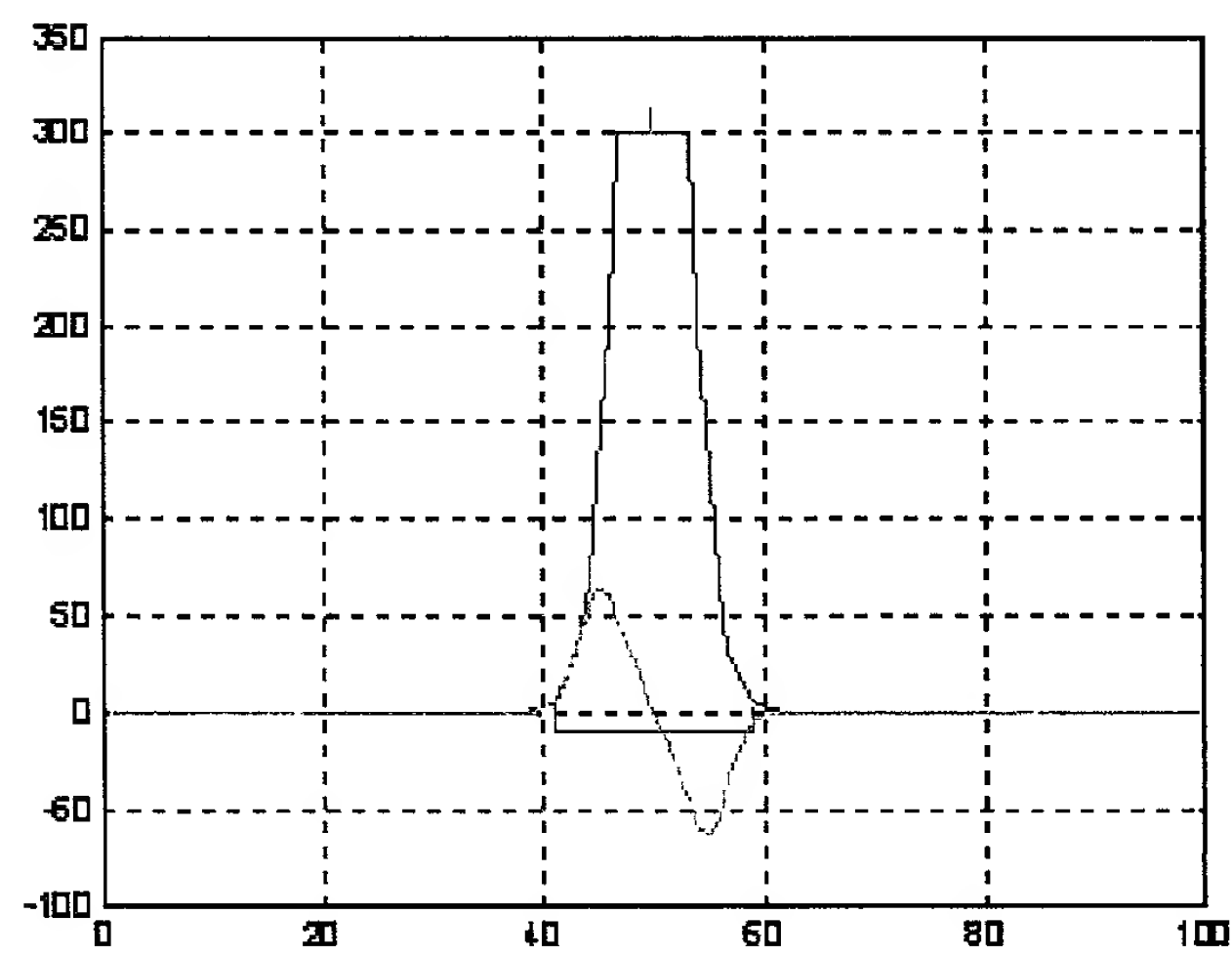


Figure 34: How to avoid the sharp corner artifact.

```

in = read.table("data/1997-2000",
               header = TRUE,
               as.is = TRUE,
               sep = ";",
               stringsAsFactors = FALSE)

# Create a vector of years
years = 1997:2000

# Create a vector of states
states = c("AL", "AK", "AZ", "CA", "CO", "CT", "DE", "FL", "GA", "HI", "IL", "IN", "IA", "KS", "KY", "LA", "MA", "MD", "ME", "MI", "MN", "MO", "MS", "MT", "NE", "NH", "NJ", "NM", "NY", "NC", "ND", "OH", "OK", "OR", "PA", "RI", "SC", "SD", "TN", "TX", "UT", "VA", "VT", "WA", "WI", "WY")

# Create a matrix of data
data = matrix(NA, nrow = length(years), ncol = length(states))

# Loop over years and states to fill the matrix
for (year in years) {
  for (state in states) {
    data[year, state] = in[paste(year, state, sep = "-"), ]
  }
}

# Print the matrix
print(data)

```

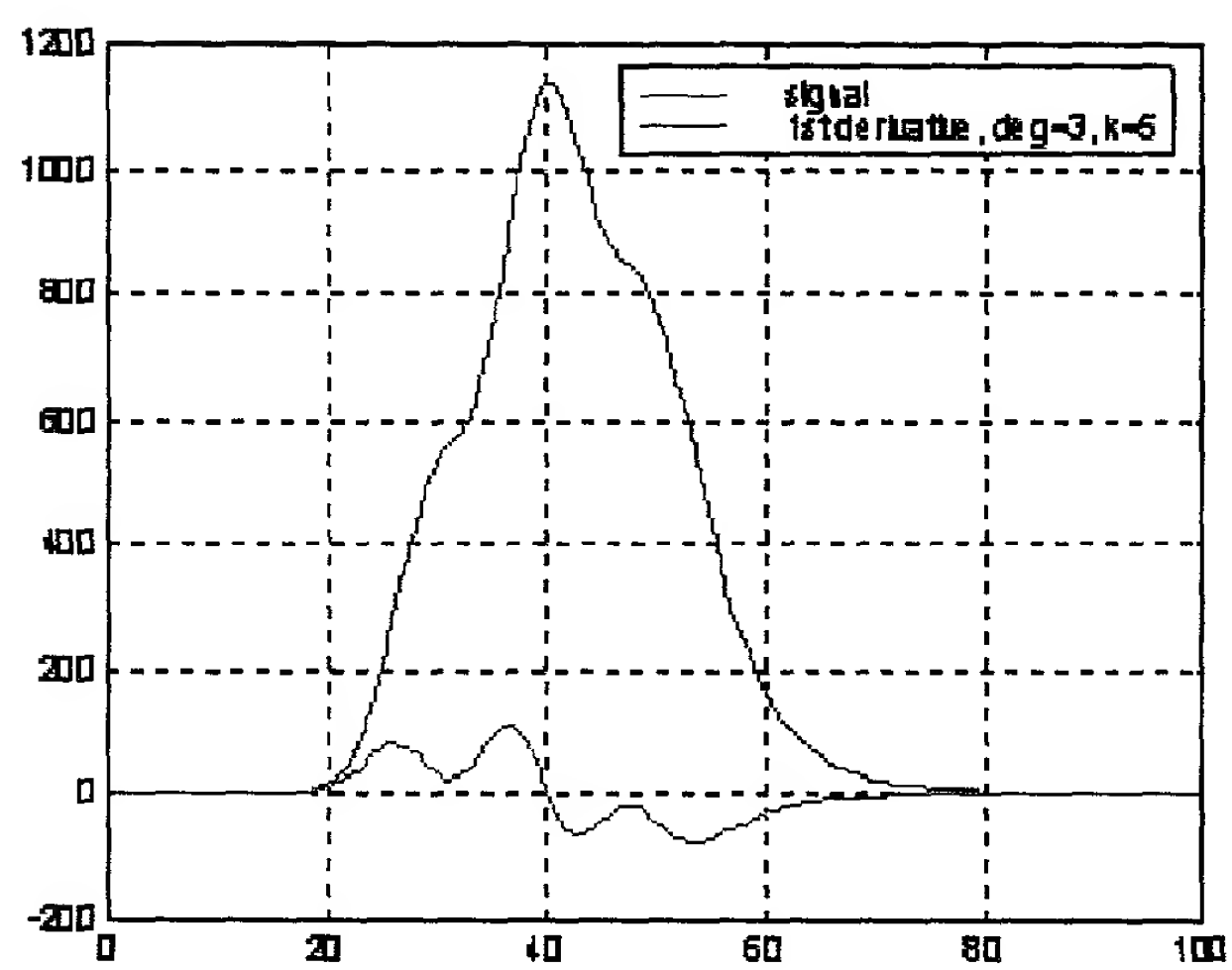


Figure 35: Peak with shoulders.

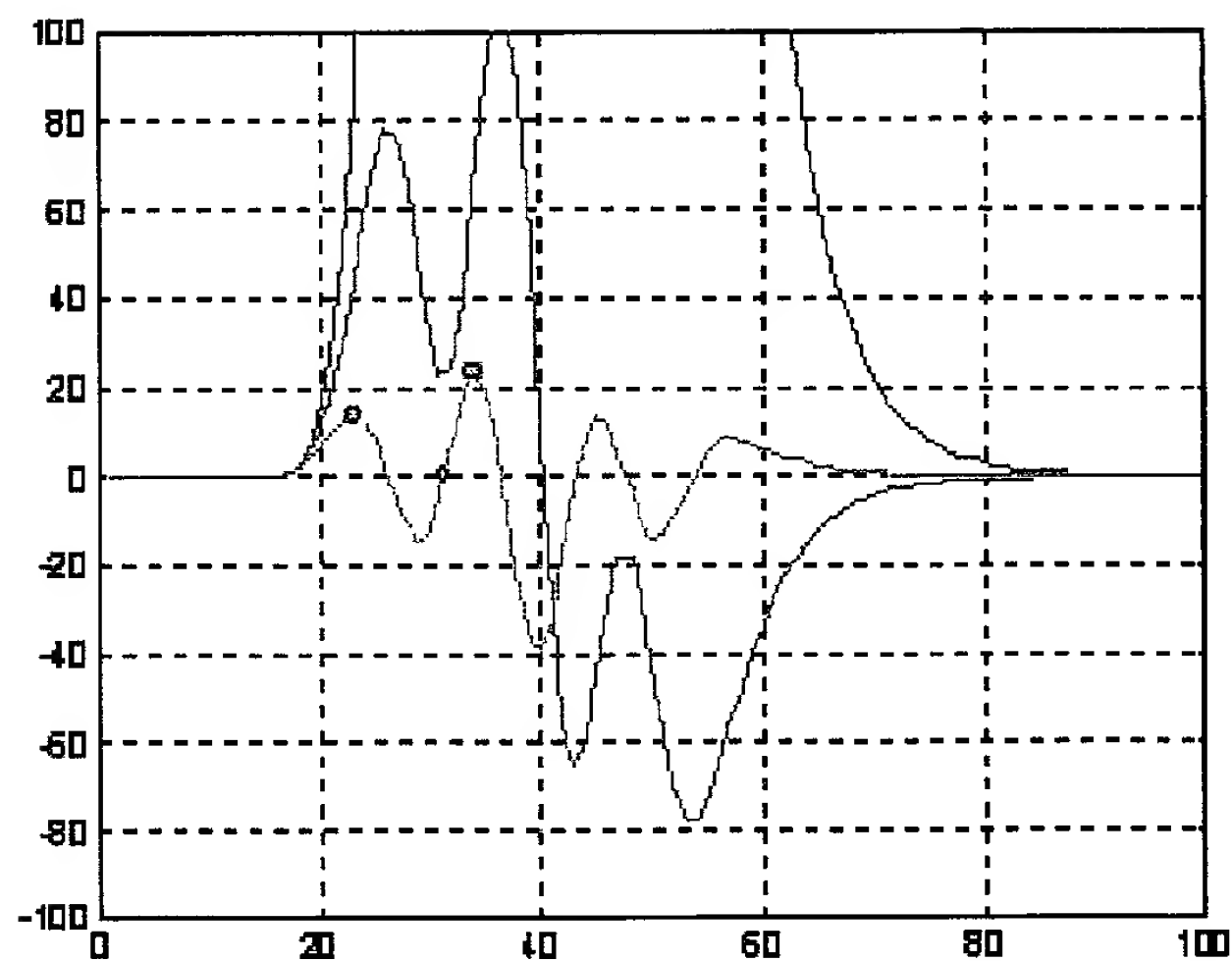


Figure 36: Finding a left-bank shoulder by analyzing the second derivative.

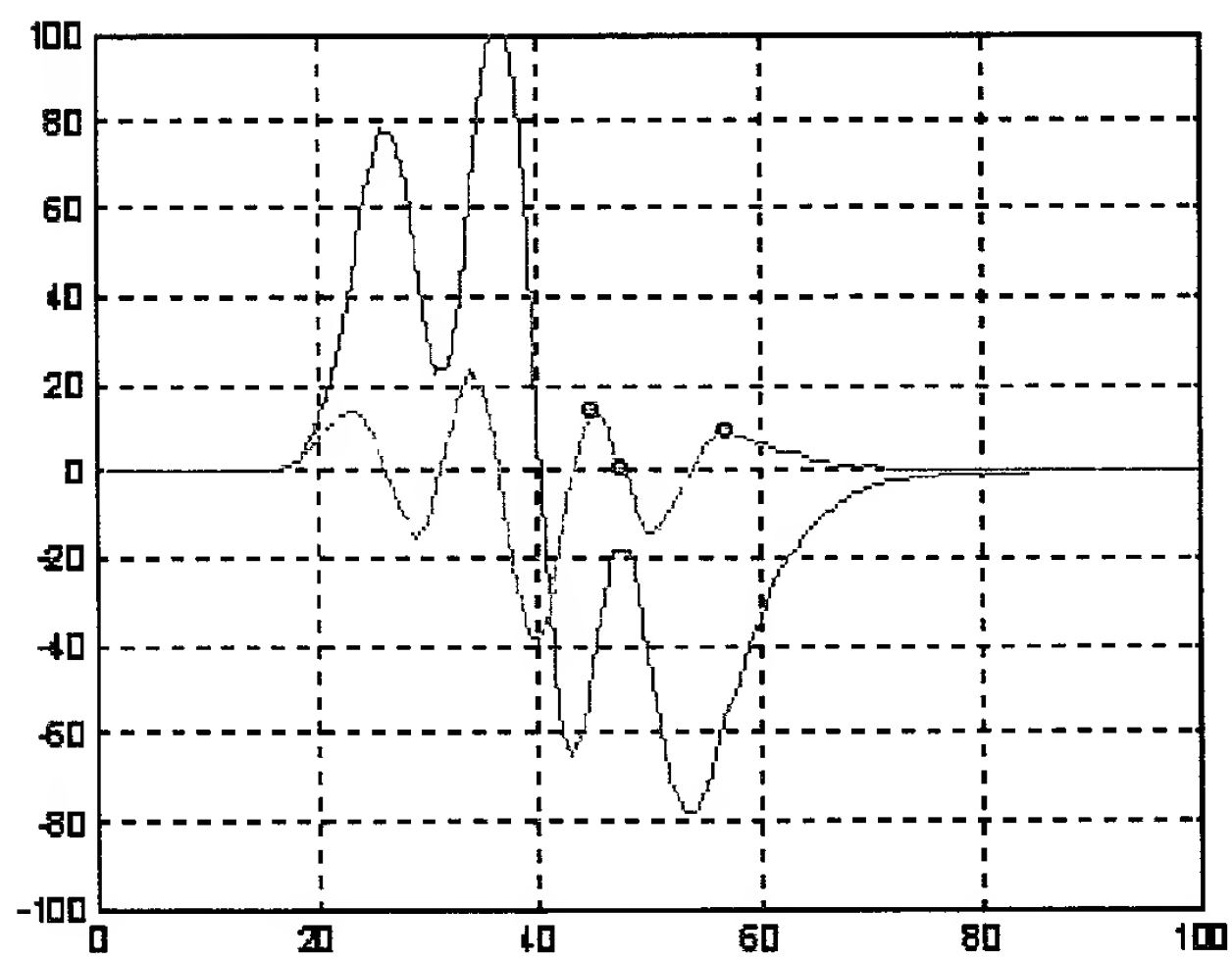


Figure 37: Finding a right-bank shoulder by analyzing the second derivative.

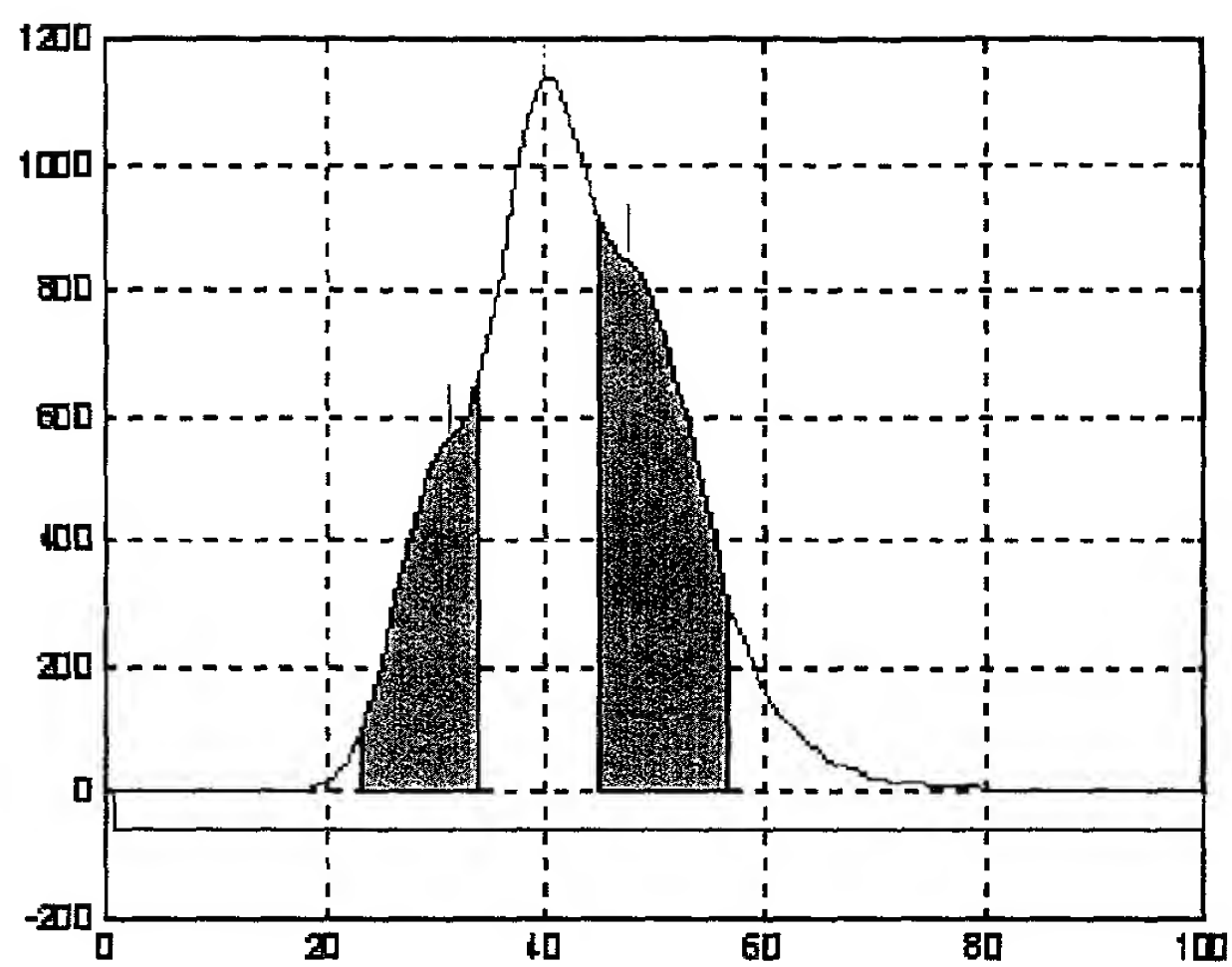


Figure 38: The complete analysis of the peak for shoulders.



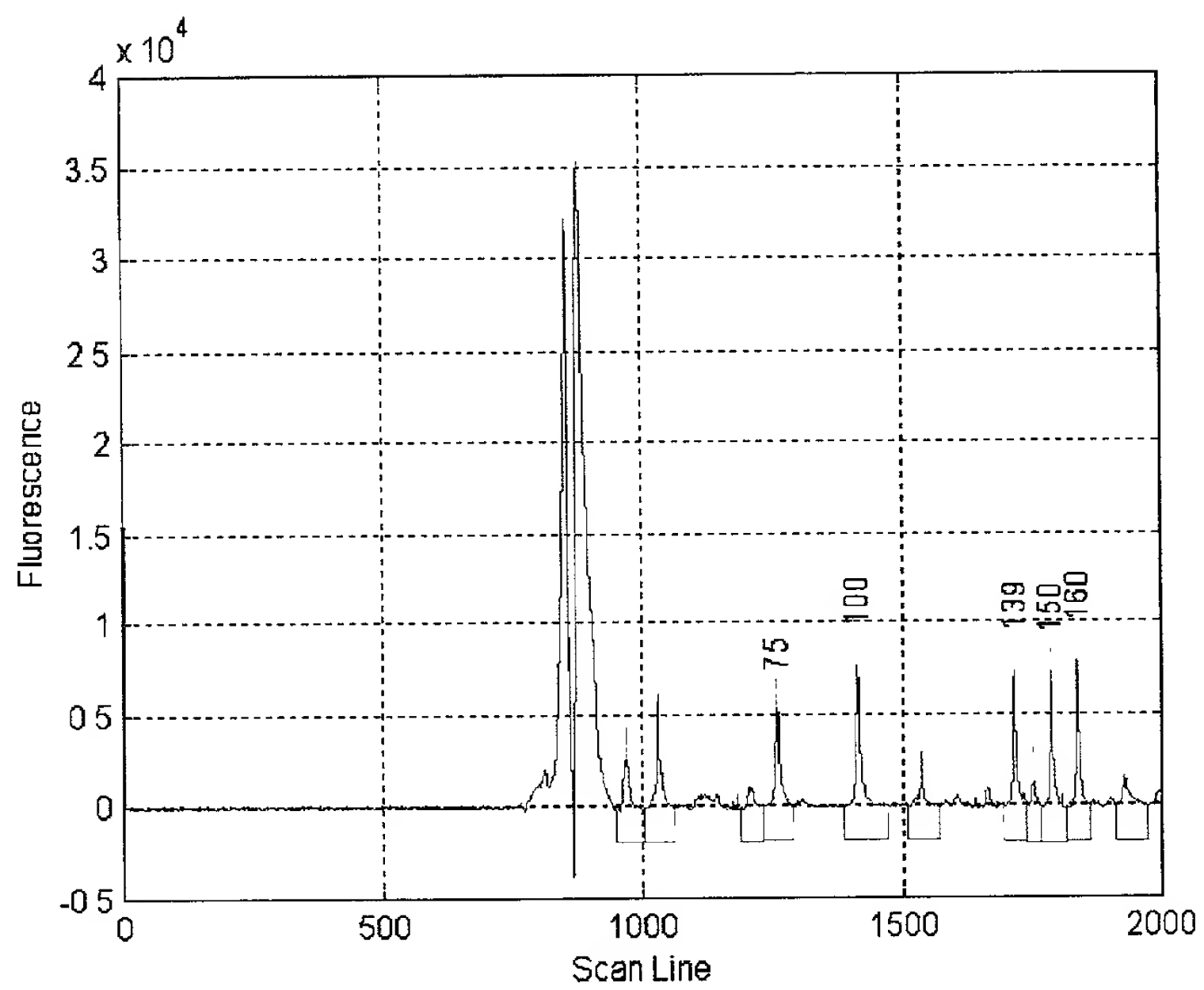


Figure 39: Example Peaks, Sizes, and a Matching.  
Peaks have red bases and green positions.

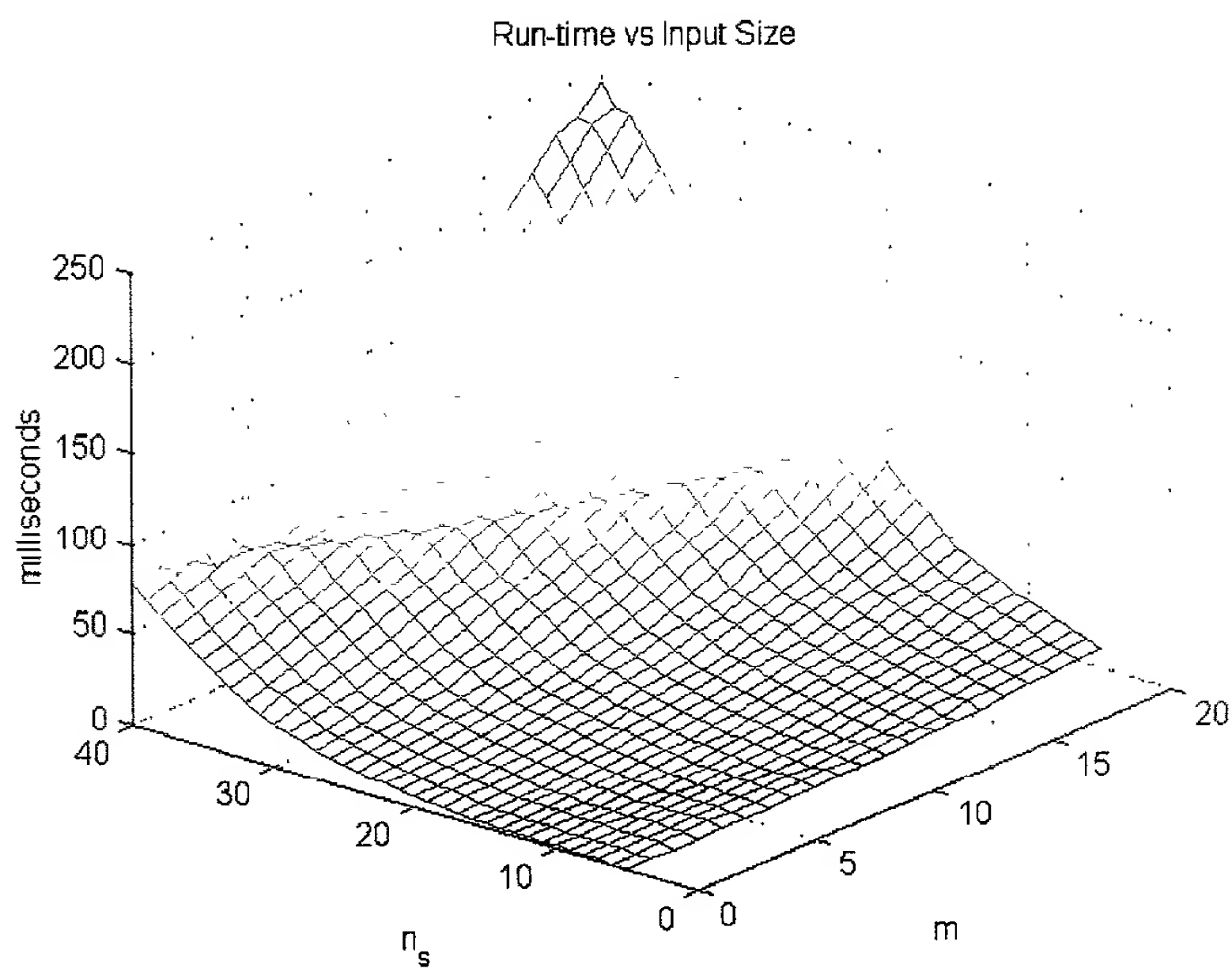


Figure 40: A mesh of execution times.

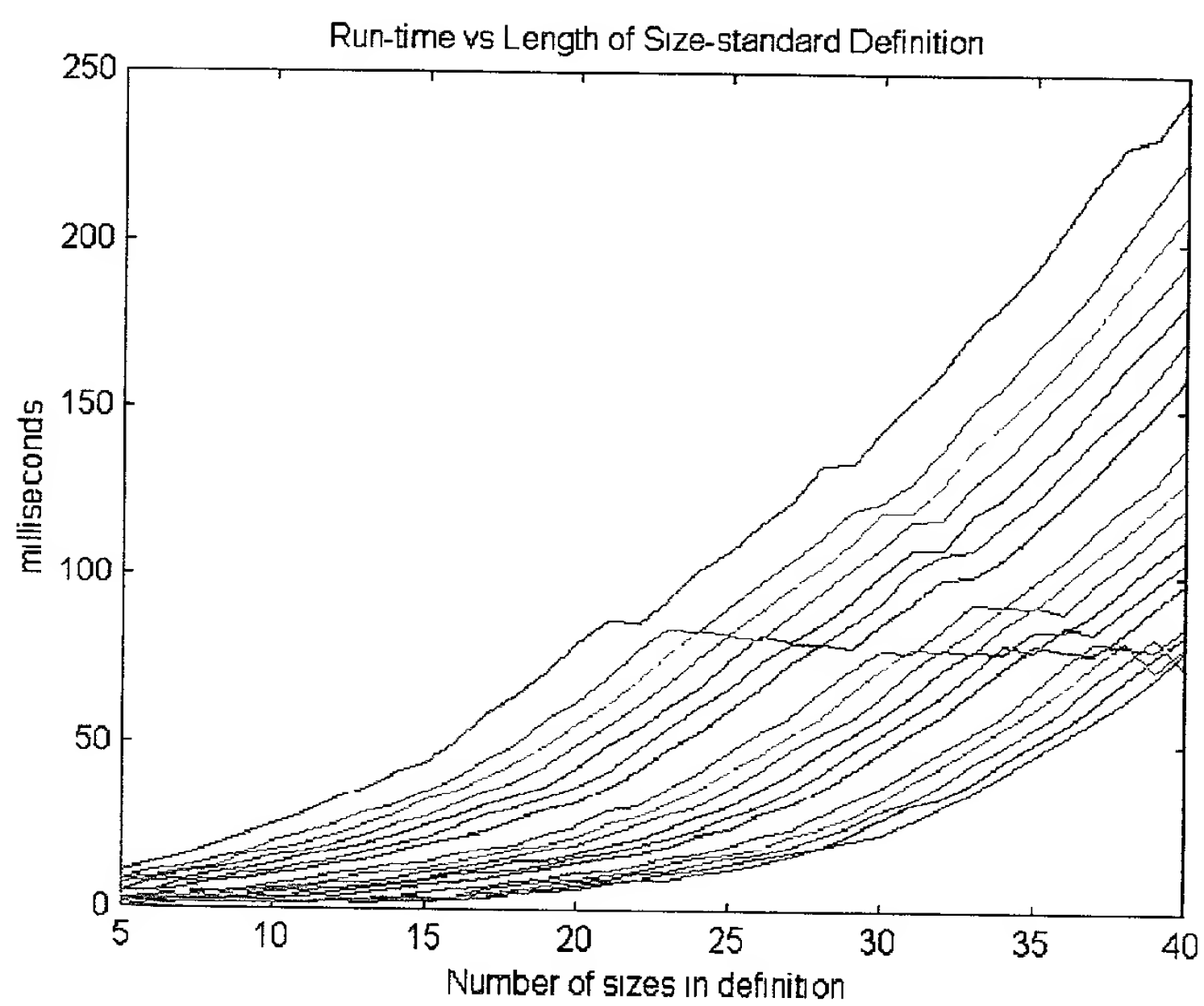


Figure 41: Each curve holds constant the number of extra peaks  $m$ .

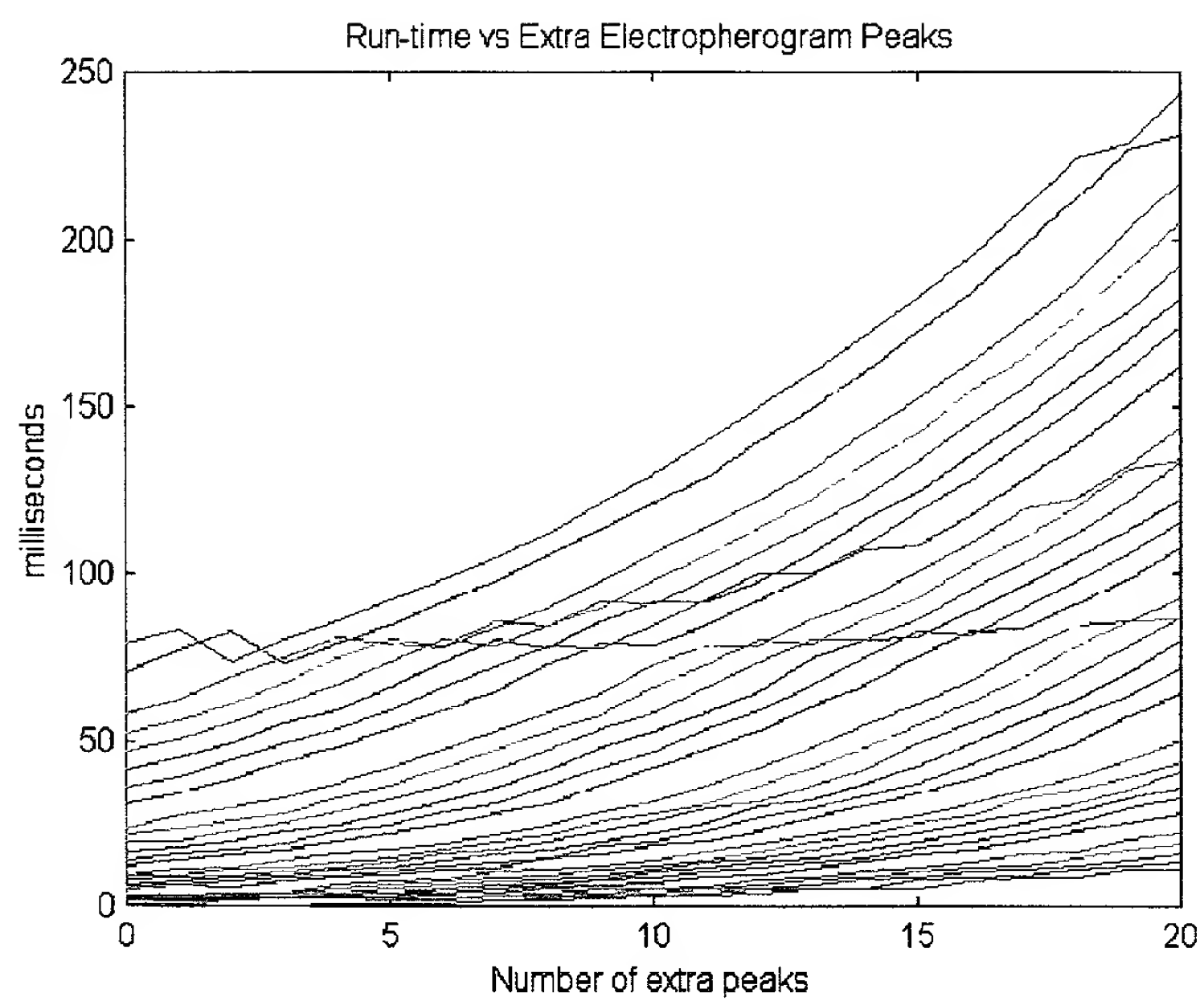


Figure 42: Each curve holds constant the number  $n_s$  of sizes in the size-standard definition.

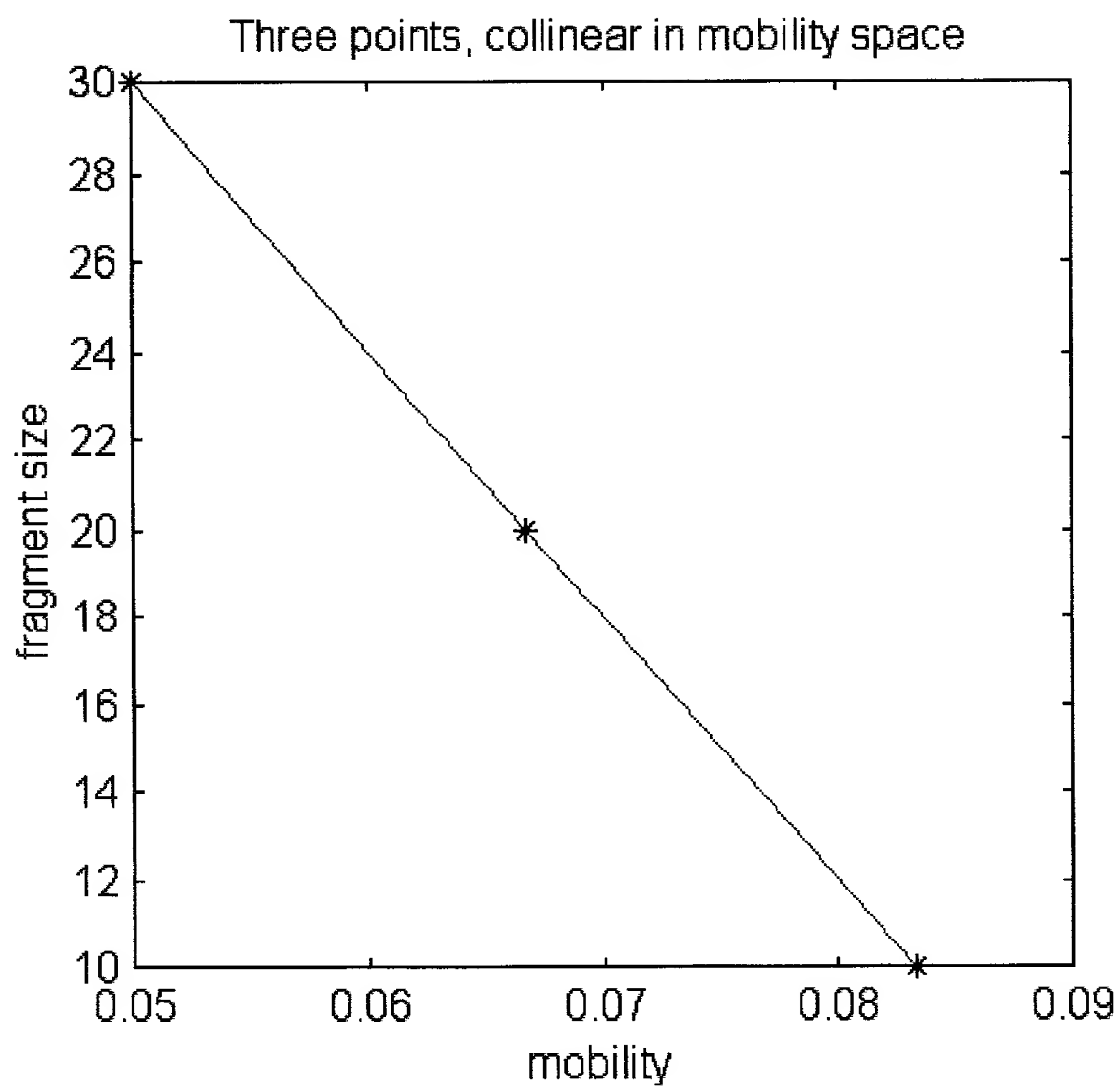


Figure 43: Three collinear points in mobility space.  
The blue line is a linear interpolation.

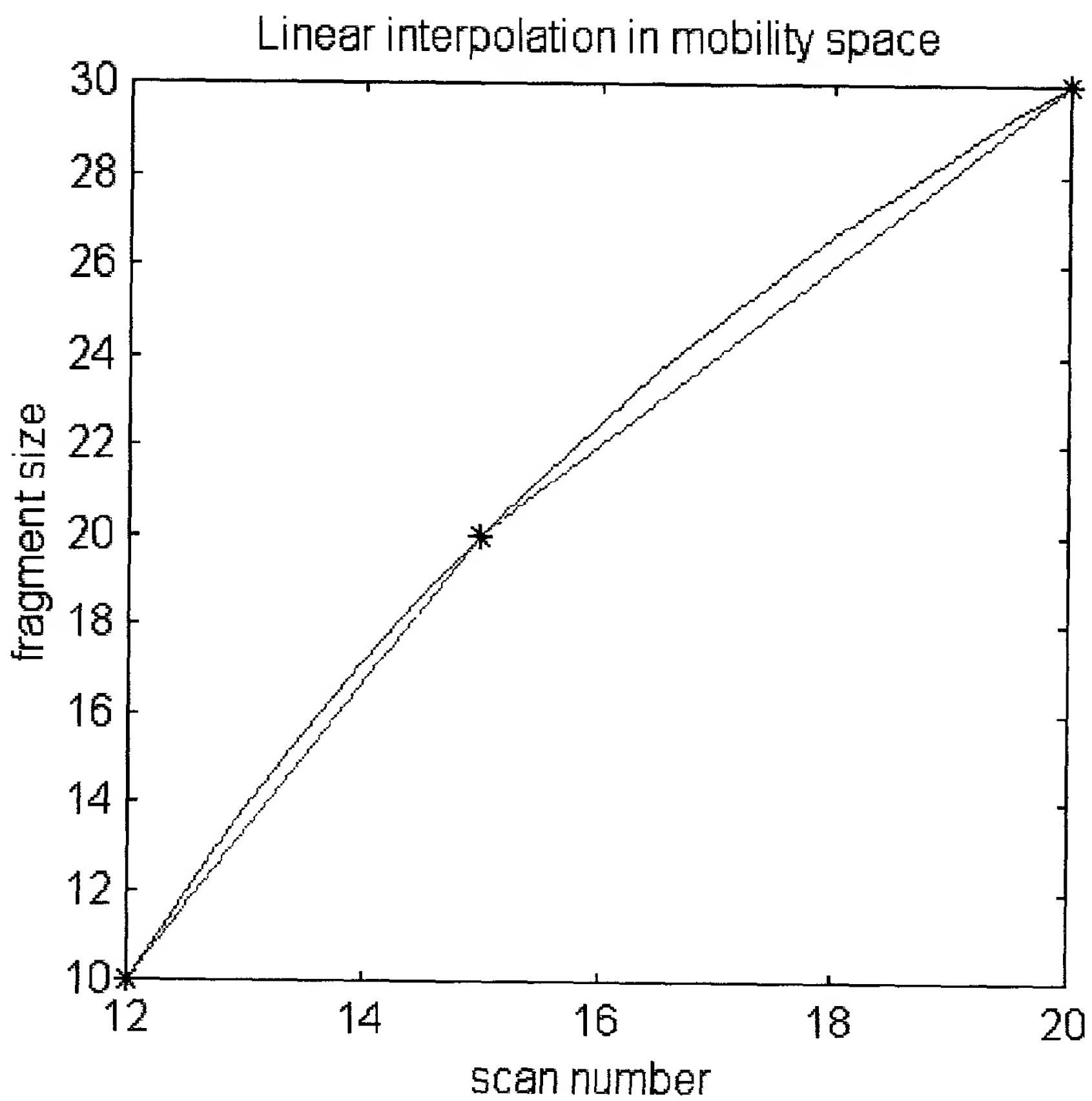


Figure 44: The three points from Figure 43 shown in size-versus-scan space. The blue line is a linear interpolation in mobility in mobility space and the red line is a linear interpolation in scan space. Note that the red line is not smooth at scan 15.

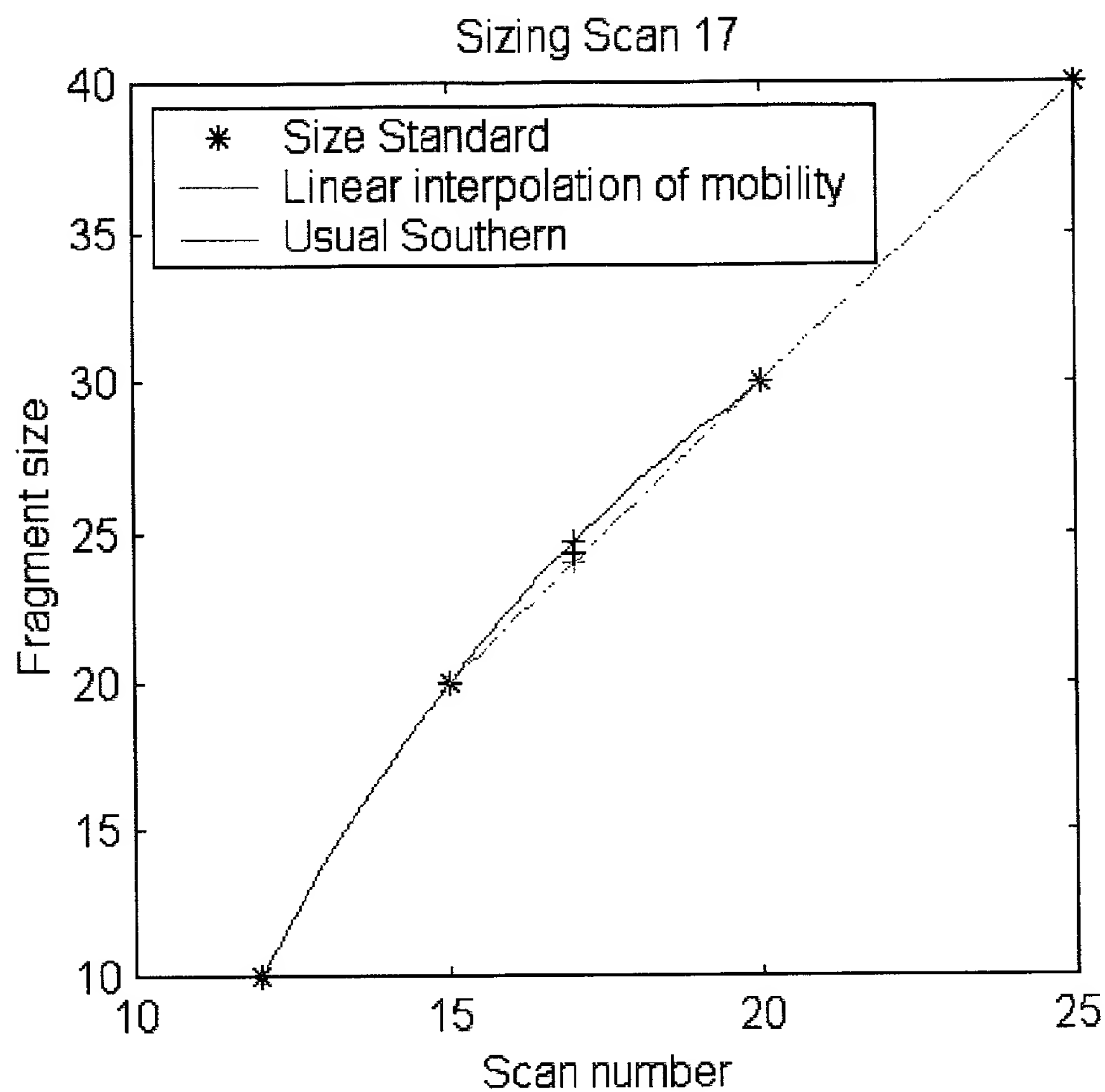


Figure 45: Three points that are collinear in size-versus-mobility space, and three that are collinear in size-versus-scan space. The blue line is a linear interpolation in mobility space and the green line is a linear interpolation in scan space. The two curves are averaged at scan 17.

Figure 46

